ESSAYS ON MERGERS, COLLUSION, AND EXCLUSIVE DEALING

by

Wei Zhao

A dissertation submitted to The Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland

October, 2013

© 2013 Wei Zhao

All Rights Reserved

Abstract

This dissertation studies antitrust issues in mergers, tacit collusion, and exclusive discount. The second chapter evaluates the welfare effects of the 1997 merger between Boeing and McDonnell Douglas in the medium-sized wide-bodied aircraft industry with a special emphasis on dynamic merger efficiencies. I develop an empirical model of multi-product firms, allowing for both learning-by-doing and product innovation in the dynamic game to quantify merger efficiency. The results show that the primary benefits from the 1997 Boeing-McDonnell Douglas merger come from accelerated learning-by-doing. Taking account of all static and dynamic effects, net consumer surplus is found to have increased by as much as $1.57 billion. In contrast, a static model ignoring dynamic learning-by-doing and innovation predicts a consumer loss of $22.53 billion due to reduced competition. These results provide important policy implications for antitrust practice by showing that ignoring dynamic effects could possibly lead to biased results and erroneous conclusions with regard to the welfare impact of a merger.

The third chapter models emergence of tacit collusion. In the context of an infinitely repeated Prisoners' Dilemma, coauthor Joseph Harrington and I provide insights demonstrating how cooperation is initiated when players cannot form explicit agreement,

but signal and coordinate through their actions. We find the longer players have gone without cooperating, the probability that they'll cooperate in future diminishes. While the probability of cooperation emerging is always positive, there is a positive probability that cooperation never occurs.

In the fourth chapter of this dissertation, I connect economic theory and antitrust practice for exclusive discount by providing a structural empirical framework that is directly applicable in antitrust practice. Numerical examples are provided to demonstrate how to apply the framework to empirically evaluate the impact of bundled or exclusive discount under different market structures. The numerical analysis also suggests that the welfare effects of bundled and exclusive discount are case specific. In addition, differences between bundle and exclusive discount are important, as it is possible for either one to hurt consumers while the other makes them better off.

Keywords:	Dynamic Merger Efficiency, Aircraft, Learning-by-doing, Innovation, Tacit Collusion, Cooperation, Bundled Discount, Exclusive Discount
JEL Classification:	C72, C73, L13, L40, L44
Advisors:	Professor Joseph E. Harrington, Jr.
	Professor Przemyslaw Jeziorski

Acknowledgements

First and foremost, I am extremely indebted to my advisors Joseph E. Harrington, Jr. and Przemyslaw Jeziorski for their continual guidance, support, and encouragement. I am grateful to Richard Spady for various insightful suggestions and C. Lanier Benkard for making the labor input data of Lockheed L-1011 available. I deeply appreciate Edmund S. Greenslet, publisher of *The Airline Monitor*, for numerous inspiring discussions about the aircraft industry and the merger. Also, I would like to thank Yonghong An, Yingyao Hu, Elena Krasnokutskaya, Yiyang "Ellen" Li, Jonathan Wright, and participants of seminars at Johns Hopkins University, SUNY Buffalo, Northeastern, and 11th Annual IIOC Conference for helpful comments, discussion, and suggestions. Finally, I would like to thank my wife Yiyang for her never-failing support and understanding, which make this dissertation complete. All remaining errors are mine.

Contents

List of Tables

List of Figures

1 Introduction

This dissertation contributes to the literature of antitrust economics. Chapters 2-4 are three separate essays on mergers, tacit collusion, and exclusive discount, respectively. The second chapter evaluates the welfare effects of the 1997 merger between Boeing and McDonnell Douglas in the medium-sized wide-bodied aircraft industry with a special emphasis on dynamic merger efficiencies. I develop an empirical model of multi-product firms, allowing for both learning-by-doing and product innovation in the dynamic game to quantify merger efficiency. Merger efficiency from learning-by-doing is then disentangled from impact of innovation and damages from reduced competition. The results show that the primary benefits from the 1997 Boeing-McDonnell Douglas merger come from accelerated learning-by-doing. Taking account of all static and dynamic effects, net consumer surplus is found to have increased by as much as $1.57 billion. In contrast, a static model ignoring dynamic learning-by-doing and innovation predicts a consumer loss of $22.53 billion due to reduced competition. These results provide important policy implications for antitrust practice by showing that ignoring dynamic effects could possibly lead to biased results and erroneous conclusions with regard to the welfare impact of a merger.

The third chapter models emergence of tacit collusion. In the context of an infinitely repeated Prisoners' Dilemma, coauthor Joseph Harrington and I provide insights demonstrating how cooperation is initiated when players cannot form explicit agreement, but signal and coordinate through their actions. There are two types of players - patient and impatient. Players are not informed of any other player's type. An impatient type is incapable of cooperative play, while if both players are patient types - and this is common knowledge - then they can cooperate. The game is thus a process of learning and revealing information on player types. Through repeated interactions, players update their beliefs about the type of the other player, which determines their own probabilities of initiating cooperative plays and revealing their types. We find the longer players have gone without cooperating, the probability that they'll cooperate in future diminishes. While the probability of cooperation emerging is always positive, there is a positive probability that cooperation never occurs.

In the fourth chapter of this dissertation, I connect economic theory and antitrust practice for exclusive discount by providing a structural empirical framework that is directly applicable in antitrust practice. In the model, I highlight the difference between product alternatives and bundle alternatives and the difference between bundle discount and exclusive discount. Numerical examples are provided to demonstrate how to apply the framework to empirically evaluate the impact of bundled or exclusive discount under different market structures. The numerical analysis also suggests that the welfare effects of bundled and exclusive discount are case specific. In addition, differences between bundle and exclusive discount are important as it is possible for either one to hurt consumers while the other makes them better off.

2 Estimating Dynamic Merger Efficiencies with an Application to the 1997 Boeing-McDonnell Douglas Merger

2.1 Introduction

> "A primary benefit of mergers to the economy is their potential to generate significant efficiencies... which may result in lower prices, improved quality, enhanced service, or new products." (2010 U.S. Horizontal Merger Guidelines)

One of the central duties of the Federal Trade Commission (FTC) and the Antitrust Division of the U. S. Department of Justice (DOJ) is to evaluate the potential impact of a merger between competing firms on the welfare of consumers. Mergers that would make consumers worse off are either to be restructured in order to avoid such detrimental effects or challenged and prevented. In light of the size and number of companies involved in merger activity, the potential welfare impact is significant. As reported in the most recent *Hart Scott Rodino Annual Report*, there were 1,450 proposed transactions involving large companies in 2011, with a total capitalization of almost one trillion dollars.

In evaluating a prospective merger, the approach is to compare the pre-merger outcome with a forecast of the post-merger outcome. This comparison typically takes the form of comparing current prices with some projected prices if the merger were to occur. To generate that projection, it is common to hold firms' costs and the quality of their products fixed and to estimate what the impact on price would be if a firm's assets were acquired by a competitor. While there are some variants to this approach, for example, it may be recognized that some products would be removed or some immediate cost reductions realized, the evaluation still takes the form of a short-run analysis. The fundamental question asked is: what will happen to consumer welfare in the short-run in response to this merger?

It is well recognized, however, that the primary efficiencies from some mergers are likely to be dynamic, as they are realized over time and are endogenous to firms' decisions in the post-merger environment. Such dynamic efficiencies can come from a reduction in cost because of learning-by-doing or altered incentives to invest in reducing marginal cost, from better products due to investment or adoption of new technologies, and from future entry and exit (perhaps involving additional mergers and acquisitions). For example, the international hard drive disc (HDD) market has experienced a series of major mergers in recent years. Maxtor and Samsung were acquired by Seagate in 2006 and 2011, respectively, and Hitachi was sold to Western Digital in 2012. The most significant impact on consumer welfare from this altered market structure may lie not with how it affects price in the short-run but rather its impact on product cost and quality in the long-run. Will firms have stronger or weaker incentives to invest and innovate? Effectively addressing such questions is central to a proper evaluation of the welfare effects of these mergers.

Though dynamic efficiencies are well-recognized as potentially substantial, they have not played a significant role in merger evaluation by antitrust authorities because of the lack of methods for estimating these efficiencies.[1] Furthermore, while we can speculate as to what the dynamic welfare effects of a past merger might be, there has been little research that actually estimates these effects. The primary objective of this paper is to contribute to studies on these policy issues by quantifying dynamic efficiencies from the 1997 Boeing-McDonnell Douglas merger. In achieving this, I develop an empirical model targeting the aircraft industry and then compare efficiencies estimated using a dynamic model to those estimated from a static model.

The empirical model encompasses two common dynamic forces relevant to industry performance and thus to the evaluation of a merger: learning-by-doing[2] and improvements in product quality[3]. These forces are encompassed in a model with multi-product firms that compete in an infinite-horizon dynamic game. In each period, a firm decides how much to produce (which may be a vector of quantities if it has multiple products), while taking into account its impact on current profit and its future profit stream through the effect of output

[1]The 2010 U.S. Horizontal Merger Guidelines indicate that dynamic efficiencies, “such as those relating to research and development, are potentially substantial but are generally less susceptible to verification.”

[2] Traditional industries benefiting from learning-by-doing include: aircraft, shipbuilding, semiconductors, fuel cell vehicles, oil drilling, photovoltaics, machine tools, metal products, nuclear power plants, and chemical processing. Recent works in estimating learning-by-doing include Benkard (2000) for aircraft, Thompson (2001) and Thompson (2007) for shipbuilding, and Gowrisankaran, Ho, and Town (2006) for surgical procedures. See Besanko, Doraszelski, Kryukov, and Satterthwaite (2010) for a complete list of learning-by-doing estimation works.

[3]Competition in quality improvements is important in high technology industries, such as biotechnology and pharmaceuticals, medical instruments, aircraft, automobiles, computer hardware and software, cell phones, and game consoles. However, there are limited intra-industry empirical studies on relationships between quality improvements and market competition. Examples include Goettler and Gordon (2011) and Nosko (2010) for the CPU market, and Hashmi and Biesebroeck (2010) for automobiles.

on the firm's experience. Experience is a state variable that rises as a firm's past output accumulates (learning), but also depreciates over time (forgetting). Learning-by-doing is modeled by having unit production cost be a decreasing function of experience. Also, a firm's production is allowed to have spillover effects with regards to experience accumulation from that firm's other products and also its competitors' products; the magnitude of these spillover effects are allowed to depend on ownership and product characteristics. In addition to deciding how much to produce each period, a firm decides whether to invest in improving the quality of its products. These potential product upgrades are exogenously generated from outside of the industry. Adoption of an upgrade incurs a direct cost but also an indirect cost through a setback in experience; for example, Levitt, List, and Syverson (2012) found for the automobile industry that "introducing a new model variant into production does cause productivity setbacks." For this setting, firms are assumed to behave according to a Markov Perfect Equilibrium in which they decide on production and upgrades in each period given the state variables of firms' experiences and product qualities, as well as given the stochastic realization of market size, product characteristics, and upgrading costs.

Before moving on to specifics relating to the aircraft industry, let us consider the possible welfare implications of a merger within this framework. A merger might hurt consumers because reduced competition creates the incentive to restrict production and raise price; this is the traditional market power effect. A merger might also generate dynamic efficiencies in several ways through its impact on the evolution of production experience. First, there is an immediate benefit in lowering marginal cost for products of the merged firm because accumulated experience is shared; This is due to possible spillover of experience across products, serving to lower cost, raise output, and elevate consumer welfare. Second, the merged firm may choose to alter its product line, for example, by shutting down some of the products of the acquired firm. Fewer products means less variety (which makes consumers worse off) but also more output per product, implying faster experience accumulation, lower unit cost, and lower future prices (which makes consumers better off). Third, future experience might be more effectively shared between different products within the same firm (within-firm spillover) than between different firms (across-firm spillover), which again will produce lower costs after the merger.

A second source of dynamic efficiency comes from altered incentives for quality improvements. The direction of this effect is ambiguous. After a merger, softened competition could discourage innovation but enlarged market share may mean a bigger benefit from a better quality product. which would stimulate incurring the fixed cost to innovate. If quality improvements negatively impact experience and raise unit cost then this will further complicate the evaluation. Assessing how these forces net out in terms of firm behavior and consumer welfare will then require estimating parameters, solving the dynamic model for equilibrium behavior, and simulating the industry path with and without a merger.

Having developed this empirical model, I then apply it to the medium-sized wide-bodied aircraft industry to evaluate the 1997 merger between Boeing and McDonnell Douglas. Prior to the merger, the market was occupied by three firms, Boeing, Airbus, and McDonnell Douglas, who were producing four products (A330, A340, B777, and MD-11) in the medium-sized wide-bodied aircraft market.[4] Immediately after the merger, the new Boeing company shut down production of MD-11. Manufacturing aircraft is labor-intensive and learning-by-doing is commonly recognized as an important feature in the industry.[5] Boeing 777 was introduced only two years before the merger, with submodels of B777 arriving soon after the merger. Thus, by ceasing production of MD-11, Boeing hoped to achieve lower marginal cost more rapidly for its B777. Besides learning-by-doing, innovation through upgrades is another distinct feature of the aircraft industry. New generations of aircraft were introduced of higher quality. This was especially so after the September 11th attacks when petroleum prices skyrocketed and airline demand for more fuel-efficient aircraft accelerated.

To evaluate the welfare effects of the 1997 Boeing-McDonnell Douglas merger, the dynamic model is solved for three different scenarios: (i) merger and the MD-11 is immediately shut down (which is what actually occurred), (ii) merger with continued operation of the MD-11, and (iii) no merger. The time series for price, consumer surplus, profit, and total surplus was computed for all scenarios. To disentangle efficiency resulting from learning-by-doing from efficiency due to quality improvements and the market power effect, I solved an

[4]The merger of the two companies affects the entire aircraft industry. However, I will focus on its impact on the medium-sized wide-bodied aircraft industry only, which can be viewed as an isolated market from other aircraft industries as discussed in Section 2.4.1.

[5]The aircraft industry is the market where learning-by-doing was first recognized. See Wright (1936).

additional model that does not allow for quality improvements and still another that does not allow for either learning-by-doing or quality improvements. The results show that the primary benefits from the 1997 Boeing-McDonnell Douglas merger come from accelerated learning-by-doing rather than from a higher rate of innovation. Furthermore, the dynamic efficiencies generated by the merger are large enough to exceed the static market power effect, which is about \$20 billion. Taking account of all static and dynamic effects, net consumer surplus is found to have increased by as much as \$1.57 billion. These results show that ignoring dynamic effects can lead to biased results and erroneous conclusions with regard to the welfare impact of a merger.

This paper is directly related to three lines of research: dynamic effects of a merger, learning-by-doing in the aircraft industry and other industries, and dynamic innovations within an industry. Gowrisankaran (1999) is one of the first papers that theoretically examined the dynamic effects of a merger. Performing numerical analysis within the Ericson-Pakes framework (Ericson and Pakes (1995)), firms were modeled as choosing investment to expand capacities dynamically, with endogenously generated mergers. Gowrisankaran (1999) assumed marginal cost is fixed and common across firms. The impact of a merger on consumer welfare was not a central issue in that paper. Chen (2009) also examined these issues theoretically and had firms make dynamic investment decisions affecting capacity accumulation, which impacted marginal cost over time. That analysis explored the bias in static merger analysis when dynamic investment is ignored. Stahl (2009) estimated cost and revenue parameters for the broadcast television industry, where costs were estimated as residuals of firm behavior unexplained by revenues. That paper focuses on the consolidation process itself rather than evaluating merger-generated efficiencies and thus does not solve the dynamic oligopoly model. Benkard, Bodoh-Creed, and Lazarev (2010) evaluated the medium- and long-run dynamic effects of airline mergers and explored the effect of mergers on market structure rather than consumer welfare. Jeziorski (2011b) and Jeziorski (2011a) studied merger impacts in the U.S. radio industry and took account of the markets being two-sided. Jeziorski (2011b) compared listeners' welfare increase from product variety with the market power effect. Jeziorski (2011a) endogenized merger decisions and found that total cost savings from mergers outweighed the loss caused by increased market power. Nocke and Whinston (2010) provided a new theoretical framework to model dynamic merger decisions where firms' choice variables other than merger decisions are assumed to be static. They derived conditions whereby an antitrust authority can maximize the present value of consumer surplus by using a myopic merger review policy. I contribute to this line of research by introducing a model that focuses on the endogenous dynamics of cost and product quality.

The empirical learning-by-doing literature encompasses a wide array of industries. This paper is most closely related to the pioneering research of Benkard (2000) and Benkard (2004). Benkard (2000) introduced the concept of forgetting to explain the rise in cost for the Lockheed L-1011, and Benkard (2004) allowed for a learning curve in a dynamic oligopoly model with four single-product firms, estimating welfare under several counterfactual scenarios with a social planner and a monopoly. This paper follows this methodological path but focuses on merger evaluation. I extend the empirical model to allow for multi-product firms, dynamic quality improvements, and the spillover effect of learning curves. In my model, merger efficiencies are likely to arise either through accumulation of experience due to combining output and the spillover effect or through a higher probability of upgrading products. Although the spillover effect of the learning curve has not been widely investigated for the aircraft industry,[6] it has been modeled and estimated for other industries, including semiconductors (Irwin and Klenow (1994)), shipbuilding (Thornton and Thompson (2001)), fuel cell vehicles (Schwoon (2008)), steel (Ohashi (2005)), and health care (Chandra and Staiger (2007)). However, those papers are not targeted at evaluating mergers in the context of a dynamic game, and none of them simultaneously estimated within-firm spillover and across-firm spillover, which could be significant factors in calculating merger efficiencies.

With respect to the empirical literature on innovation, this paper is most closely related to Goettler and Gordon (2011), which examines the microprocessor industry. Both Goettler and Gordon (2011) and my work use the concept of a product's quality *relative* to an outside good whose quality is changing over time; this is a modeling device first proposed in Pakes

[6]Benkard (2000) modeled a submodel spillover effect among submodels of an aircraft type but no cross-product or cross-firm spillover effect and a complete within-model spillover is assumed in Benkard (2004). The international trade literature studies knowledge-spillover in the sense of technology transfer across countries and across industries. See Grossman and Helpman (1995) for a review of that literature and Niosi and Zhegu (2010) for a review of the aircraft industry specifically.

and McGuire (1994). One major difference of my paper from that by Goettler and Gordon (2011) is in the modeling of quality evolution of the outside good. Goettler and Gordon (2011) fixed the difference in quality between the industry frontier product and the outside good. Thus, in their model, the outside good upgrades automatically when the product with the highest quality upgrades. In contrast, I let the quality upgrade of the outside good be exogenous and evolve stochastically. Although it is appealing to endogenize outside good evolution in a single-product firm model as in Goettler and Gordon (2011), their method might not be suitable for multi-product aircraft manufacturers. Aircraft upgrades involve inventions of new patented technology that are more likely to be shared within a firm. Thus, it is less realistic to assume that when the frontier product upgrades, the good outside the market receives the same technology and upgrades automatically while a same-firm product does not. In addition, Goettler and Gordon (2011) focused on dynamic demand while fixing marginal cost for any given relative quality, while I assume static demand and concentrate on cost structure evolution.

In conclusion, I evaluate the 1997 Boeing-McDonnell Douglas merger in terms of its impact on consumer welfare by constructing a dynamic oligopoly model for multi-product firms with learning-by-doing and endogenous quality improvements. Allowing for both the evolution of cost and product quality is new to the literature on dynamic oligopoly equilibrium models. The rest of the paper is organized as follows. Section 2 presents a global view of the structural model. The data used for estimation and calibration is reviewed in Section 3. Section 4 applies the structural model to the aircraft industry. Using equilibrium strategies solved from the dynamic model, merger evaluation is conducted in Section 5. Section 6 concludes the paper.

2.2 Model

This section describes the general dynamic framework that is the basis for the model to be estimated for the aircraft industry. In describing the framework prior to putting forth the empirical model, the intent is to give readers a global view of the decisions made by firms and consumers and how the environment evolves. Then, in Section 4, this framework is populated with the specific structure that will then be estimated.

The model has multi-product firms with differentiated products that compete in both quantities and qualities. Quantity choices affect dynamic market cost structure through the mechanism of learning-by-doing while qualities are improved through innovation decisions to replace old generations of products with the next generation of higher quality products (which are exogenously produced). Thus, improvements in product quality are realized by a generation upgrade. The model is applicable to many industries for which learning-by-doing and innovation are important, including high technology manufacturing industries such as aircraft, computer hardware, tablet, and smart phone.

The industry is composed of I multi-product firms competing in discrete time over an infinite horizon. Firm $i \in \mathbb{I} = \{1, ..., I\}$ has a product set $\mathbb{J}_i$ and $\mathbb{J}$ is the union of $\mathbb{J}_i$ for all i. Size of $\mathbb{J}_i$ and $\mathbb{J}$, denoted by J_i and J, are thus number of products in firm i and in the industry, respectively. Exit and entry on both firm and product level are assumed away.[7] However, they can be easily incorporated in the model.[8] Quantity of product j from firm i at period t is denoted as $q_{i,j,t}$, or simply as $q_{j,t}$ when there is no need to specify to which firm the product belongs.

In the remainder of this section I discuss modeling of the demand function and production cost function to be used when firms are making dynamic decisions. Then, I introduce structures on generation upgrade decisions. The section is concluded with a description of the dynamic game.

2.2.1 Demand Function

Demand is determined by both the market size M that follows an exogenous stochastic process and characteristics of all products in the market. Characteristics of product j are classified into 3 categories. X_j represents all fixed characteristics of product j. G_j is the relative generation of product j, which measures product quality and evolves following endogenous innovation decisions. (I will explain it more fully later.) Finally, ξ_j captures characteristics unobserved to econometricians that evolve exogenously, such as product suitability. Let X, G and ξ denotes the vector of X_j, G_j and ξ_j, respectively, of all products.

[7]See section 2.4.5.3 for a more detailed discussion on exit and entry.
[8]See Doraszelski and Pakes (2007) for an example of modeling exit and entry.

It is assumed that consumers do not engage in intertemporal substitutions. Their choices of demand are solely based on current period product characteristics. Therefore, I assume that when (X, G, ξ) and an industry quantity vector Q is given in a period, the inverse demand function $P = P(Q; X, G, \xi, M)$ is single valued and taken as given for firms.

2.2.2 Production Cost Function

Production cost of product j in period t, $C_{j,t}$, is a function of quantity $q_{j,t}$ and experience level $E_{j,t}$. $C_{j,t}$ is assumed to be increasing in $q_{j,t}$ and decreasing in $E_{j,t}$. Thus, experience helps to lower production cost. $E_{j,t}$ itself is a function of the experience level from last period $E_{j,t-1}$ and the quantity vector of last period Q_{t-1}. I introduce $E_{j,t}$ so that instead of tracking the entire product history, I can just use $E_{j,t}$ as a state variable in the dynamic game. I restrict $E_{j,t}$ to be increasing in both $E_{j,t-1}$ and any $q_{k,t-1}$, $\forall k \in \mathbb{J}$. This implies that experience accumulates over time both through direct learning from production ($q_{j,t}$) and spillover from production of other goods ($q_{k,t}$, $k \neq j$). Forgetting is incorporated in the model in the form of depreciation of experience E_j as $\frac{\partial E_{j,t}}{\partial E_{j,t-1}} < 1$.

2.2.3 Generation Upgrade

I assume that product innovation can be characterized into discrete generations, with higher generations providing higher utility for consumers.[9] For an industry with everlasting innovations and infinite horizon, it is natural to believe that each product has infinitely many generations $g_j \in \{1, 2, 3, ...\}$. However, since the generation of each product is going to be a state variable in the dynamic game, direct modeling of $g_j \in \{1, 2, 3, ...\}$ will explode the state space and make it empirically intractable. Also, it is too restrictive to assume that there is some maximal level of generation. Therefore, to deal with this dimensionality issue, quality is measured as quality relative to an outside good, where the outside good stochastically improves over time, and the difference in quality between a firms' product and the outside good is bounded. Formally, relative quality is defined as

$$G_{j,t} = g_{j,t} - g_{0,t}$$

[9] See Section 2.4.4.1 for definition and reasoning for generation upgrades in the aircraft industry.

where $g_{0,t}$ is the generation level of the outside good. Relative generation of all products G_t is assumed to contain all of the information of g_t that is relevant in determining the demand function.

The model then tracks relative generations instead of absolute ones.[10] This modeling method helps to solve the dimensionality problem for industries where, given an appropriate definition of generation, maximum relative generation is observed to be small. One example of such an industry is the video game console market. A generation of the game console is commonly defined by processor word-length (number of bits), and there has been hardly more than one generation gap between actively produced game consoles at any time in the history of the industry.[11] Note that by treating G_j as a product characteristic, the assumption that relative generation is sufficient in determining demand is consistent with the discrete choice model of the demand system that is widely employed in the literature. Thus, employing relative generation creates no loss of useful information in determining demand.

I assume that $g_{0,t}$ advances each period with probability p^G.[12] In the equilibrium, p^G determines the long-run industry innovation rate.

$$g_{0,t} = \begin{cases} g_{0,t-1} + 1 & \text{with probability } p^G \\ g_{0,t-1} & \text{with probability } 1 - p^G \end{cases} \tag{1}$$

Evolution of $g_{j,t}$ is controlled by joint upgrading decisions over all products of firm i owning product j, denoted as $U_i \in \{0,1\}^{J_i}$. In each period, U_i is chosen to maximize total expected value of the firm upon observing realization of a vector of random upgrading cost C_i^G for all the products firm i owns. Let $u_{j,t} \in \{0,1\}$ be the indicator of product j generation upgrading in period t as a result of joint upgrading decisions, and let $c_{j,t}^G$ be the realized upgrading cost for product j in period t. The impact of $u_{j,t}$ can be summarized by the

[10]Given the assumption that only relative generation matters, G_j can always be normalized by subtracting it from its observed mean.

[11]See a table of generations of game consoles in Liu (2010).

[12]If enough generation upgrading decisions at each state are observed, it would be better to let p^G depend on the current state of G_j for all j in order to endogenize outside good evolution.

following equation.

$$u_{j,t} = \begin{cases} 1 \rightarrow & \text{pays } c^G_{j,t};\ g_{j,t} = g_{j,t-1} + 1;\ E_{j,t} = \psi(E_{j,t-1}) \\ 0 \rightarrow & \text{pays } 0;\ g_{j,t} = g_{j,t-1};\ E_{j,t} = E_{j,t-1} \end{cases} \tag{2}$$

where $\psi(x)$ is a given function, with the property $\psi(x) < x,\ \forall\ x$, that models setback in experience level when upgrading a product. Thus, when product j is upgraded in period t, its generation will increase by 1 while incurring an upgrade cost of $c^G_{j,t}$ and a setback in experience to $\psi(E_{j,t-1})$.

2.2.4 Dynamic Game

For the dynamic game, each product has three states variables: experience level E_j, relative generation G_j, and unobserved characteristics ξ_j. The state of the industry is then characterized by a state profile $\omega = (E, G, \xi, M)$, where M is the overall market size. Firm i makes joint decisions in upgrading all its products, U_i, and in quantity choices of those products, Q_i. Each period in the game can be divided into three stages as follows:

- **(i) Nature Stage**
  - Nature draws shocks on demand (M and ξ) and innovation of the outside good ($g_{0,t}$). All draws are immediately observed by all firms.
- **(ii) Innovation/Upgrading Stage**
  - **(ii.a)** Firms learn their upgrading cost, which is private information.
  - **(ii.b)** Firms simultaneously make adoption decisions (U_i). Resulting new generation levels of all products are immediately observed by all firms.
- **(iii) Production and Learning Stage**
  - Firms compete in a simultaneous quantity competition game. Experience level for each product is realized based on quantity choices and is revealed to all firms.

Note that experience state evolves in both stage (ii) and (iii), while generation state changes in stage (i) and (ii). Quantity and upgrading decisions are made in different stages.

Thus, expected future values need to be constructed differently when solving for optimal quantity and upgrading policies. To deal with these complexity, I found it very helpful to be specific about stages for ω. Hereafter, I will denote state profile at the beginning of Stage (ii) as ω and the state profile at the beginning of Stage (iii) as $\widetilde{\omega}$.

For Stage (ii), since firms do not observe other firms' realized upgrading costs and upgrading choices when making their own decisions, they have to put probabilities

$$Pr_k^{\omega} = \text{Probability of choosing } U_k^{\omega}$$

on competitor k's possible moves. In the following discussion on solving for Pr_k^{ω}, I drop superscript ω on U_i for simplicity and all the discussions are with respect to a given state profile ω. Denote firm i's expected value, excluding upgrading cost, of choosing U_i as $EV_i^{U_i}$. $EV_i^{U_i}$ is the summation of expected values across all products firm i owns and the expectation is over other firms upgrading probabilities Pr_k^{ω}. Let U_i and U_i' be two different vectors of choices from the set $\{0,1\}^{J_i}$. The vector U_i will be chosen if it gives firm i the largest net continuation value (expected future value less upgrade cost). Thus, the probability of choosing the vector U_i is simply given by the probability of net continuation value with respect to U_i exceeding that with respect to any other choice vector U_i', i.e.

$$Pr_i^{U_i} = Prob[(EV_i^{U_i} - C_i^G \cdot U_i) \geq (EV_i^{U_i'} - C_i^{G\prime} \cdot U_i'), \forall\, U_i' \neq U_i] \tag{3}$$

Note that by allowing firms to have multiple products, complications arise in that I need to solve for joint probabilities for each firm, which may have multiple solutions. Fortunately, introducing randomness in a separable form through upgrade cost guarantees a unique solution that can be easily solved for Equation (3). The crucial point is that given U_i, $EV_i^{U_i}$ is not a function of any c_j. The proof can be found in the Appendix.

With equilibrium Pr_i^{ω} solved from Equation (3), I now turn to equilibrium quantity choices. Since production affects future variable cost through its direct impact on experience accumulation, production decisions for each period could no longer be modeled as static. Quantities enter both the current profit function and the next period value function in the

Bellman equation. Aside from this quantity effect on future costs, the per period game is a quantity competition with heterogeneous goods and multi-product firms. The per period payoff (profit) function for product j is

$$\pi_j^{\widetilde{\omega}} = p_j(Q; X, G, \xi, M)q_j - C_j(q_j; E_j). \tag{4}$$

Let ρ denote the discount factor. Joint optimal quantity policies for firm i are solved from:

$$\max_{q_j \ \forall \ j \in \mathbb{J}_i} \left(\sum_{j \in \mathbb{J}_i} \pi_j^{\widetilde{\omega}} + \rho E[V_j(\widetilde{\omega}'|\widetilde{\omega}, Q)] \right)$$

where next period values are in prime terms. Value function for product j, denoted as $V_j^{\widetilde{\omega}}$, is then defined by the Bellman equation:

$$V_j^{\widetilde{\omega}} = \pi_j^{\widetilde{\omega}*} + \rho E[V_j(\widetilde{\omega}'|\widetilde{\omega}, Q^*)] \tag{5}$$

where "$*$" denotes value based on optimal quantity choices. The transition matrix for calculating $E[V_j(\widetilde{\omega}'|\widetilde{\omega}, Q)]$ is left in the Appendix.

In solving the model numerically, I track Pr_i^{ω} for each state profile ω and $q_j^{\widetilde{\omega}}$ and $V_j^{\widetilde{\omega}}$ for each state profile $\widetilde{\omega}$. Note that I utilize the differentiation of ω and $\widetilde{\omega}$ here. I find that tracking $V_j^{\widetilde{\omega}}$ instead of V_j^{ω} makes computation much easier.

2.3 Data

Data utilized in this paper is taken from several sources. To obtain evidence on defining medium-sized wide-bodied aircraft as a single market, I utilized route-level aircraft type and traffic data from the Bureau of Transportation Statistics. This data reveals whether medium-sized wide-bodied aircraft are mainly competing with each other on the routes they fly.

In the application to the aircraft industry, I use a nested-logit discrete choice model for the demand function. Annual fleet and deliveries data from the *Airline Monitor* are taken to construct quantities for each aircraft type each year. Annual average aircraft value data for each type is provided by *Avmark* and is used as plane prices. Market size is approximated by the total number of used and new wide-bodied aircraft using data from the *Airline Monitor*. This choice of approximation is based on the resale and rental market assumption discussed in Section 2.4.2 below. In the discrete choice model, the aircraft are heterogeneous in characteristics, and the characteristics are collected from the official websites of Boeing and Airbus, as well as various online sources. Characteristics include number of seats, maximum range, number of engines, fuselage, empty operating weight, and first flight year.

Prices need to be instrumented in the demand estimation since they are likely to be correlated with unobserved aircraft characteristics, which is the error term in the regression. Assuming that observed characteristics are uncorrelated with the unobserved components, characteristics are taken as one set of instruments. Cost shifters that are assumed to be correlated with price but not with unobserved characteristics are taken as another set of instruments. Cost shifters used include present and lagged terms of U.S. manufacturing wage rates from the*Bureau of Labor Statistics*, and aluminium prices from IMF's *International Financial Statistics Online Database.*

Production cost estimation is decomposed into three steps. First, I estimate labor input as a function of the production rate and experience. I utilize the data on direct man hours incurred by Lockheed in the production of each L-1011 aircraft for labor input;[13] The *Jet Airliner Production List* provides the first flight date of every wide-bodied aircraft produced, which is taken as the date of production.[14] Production rates and experience are constructed using quantity data and date of production. Second, the relationship between total variable costs and labor input is estimated also using data for the L-1011 program taken from Benkard (2004). Third, maintenance costs of the L-1011 plants reported in Lockheed's annual reports are used to estimate fixed costs.

In labor input estimation, quantities are likely to be correlated with unobserved productivity. Thus, I instrument quantities using a set of cost and demand shifters that are assumed to be correlated with quantities but not with unobserved productivity. Cost shifters are identical to those used in demand estimation. Demand shifters include present and lagged

[13] I am grateful to C. Lanier Benkard for making this data available.

[14] The *Jet Airliner Production List* also has ownership history of all wide-bodied aircraft, which can be used to calculate the rate of aircraft resale and rental.

terms of world and regional GDP from IMF's *International Financial Statistics Online Database* and oil price data from the *Energy Information Administration*.

Generation upgrade-related parameters are calibrated based on data from several difference sources. Fuel efficiency data from the *Airline Monitor* and operating cost difference claims reported in Boeing and Airbus newsletters are used to determine generations of aircraft. Given the definition of generation, average time before generation upgrade can be calculated using differences in first flight year across generations, which was obtained from the *Jet Airliner Production List*. Upgrading probability is then the inverse of this average time. Since generation is included as a characteristic in demand estimation, generation gap is directly obtained from demand estimation. Finally, generation upgrade costs for various aircraft models were collected from news clippings.

2.4 Empirical Application

In this section, I apply the model in Section 2.2 to the medium-sized wide-bodied aircraft industry. Depending on industry specifics and data availability, demand and cost functions described in Section 2.2 are parameterized, and parameters in the model can be estimated following at least two different approaches. First, parameters can be estimated directly in the dynamic game. The common solution method is to first build a likelihood function or moment conditions as functions of the parameters based on the data. Typical examples of moments include average firm choices (price, investment, exit and entry, etc.) and covariances between a firm's choices and a firm's own states or rival firms' states. Then one solves a constrained maximization or minimization problem with respect to the likelihood function or moment conditions by treating equilibrium conditions (Equations (3) and (5)) as constraints. Thus, when the optimization problem is solved, optimal parameter values are found together with the corresponding equilibrium of the dynamic game. Second, demand and cost parameters can also be estimated separately in a first stage, and one assumes the structures generating the estimates are unchanged in the dynamic model. The estimates are then taken as primitives in solving for the equilibrium of the dynamic game. This latter approach is computationally less burdensome than the first approach since the dynamic game only needs to be solved once, and there is no parameter searching in solving the dynamic game. However, it also requires more structure assumptions as discussed above.

Data availability can be a factor determining which approach is used. When observations are serially correlated, the entire time series of a variable, for example the price of a product, is just one observation of its evolution, which is affected by various shocks. Thus, if there is only one market in the industry, as in the case of the aircraft industry being studied here, there is just one observation for each variable to construct moment conditions or the likelihood function. This limits both credibility and the number of moment conditions that can be constructed. Hence, for this paper, I chose the second approach to evaluate the 1997 Boeing-McDonnell Douglas merger in the aircraft industry.[15]

For the rest of this section, some background information is provided regarding market definition. Then, I present the specific empirical model of the demand and cost function for the aircraft industry and discuss the estimated parameters. With demand and cost structure introduced, I turn to discussions of definition and calibration of generation upgrade. I finish this section with further analysis on applying the dynamic game to the medium-sized wide-bodied aircraft industry.

2.4.1 Medium-Sized Wide-Bodied Aircraft as an Industry

A wide-bodied aircraft is a large jet airliner with two passenger aisles. (See Figure 1 for interior arrangements of a typical 3-class-configuration wide-bodied aircraft.) Following the introduction of the first wide-bodied aircraft, Boeing 747, in 1969, only four firms were active in the industry. Of these four firms, Lockheed left the market in 1984. Nine wide-bodied types were in production during the 1990-2010 period, yet they were not all directly competing with each other due to differences in plane size and maximum flying range. Figure 2 suggests that in terms of size and range, these nine aircraft types are clustered into three groups: small (around 250 seats), medium (around 300 seats), and large (around 450 seats). The horizontal line in the figure marks the nautical distance between Beijing and New York, and is used as a benchmark separating transatlantic and transpacific routes. Differences in length of routes are continuous so this benchmark should only be viewed as a

[15] For more localized industries containing many geographic markets, the former approach might be more attractive.

guideline rather than a strict rule. However, we can see that, compared with small aircraft, medium and large aircraft have longer range and are more suitable for transpacific routes. The primary impact of the merger on market structure was the elimination of McDonnell Douglas, whose only wide-bodied aircraft then in production was MD-11. Thus, I focus on a sub-market of aircraft that directly competed with MD-11. That is, the medium-sized group, which includes A330, A340, B777, and MD-11.

Other than those nine current types shown in Figure 2, Boeing introduced B787 in 2011 as a replacement of B777 and Airbus answered with A350, an upgrade of A330, that is projected to enter the market in 2014. I treat B777 and A350 as new generation upgrades of B777 and A330 respectively in the model. In this sense, there are more than one aircraft model numbers (e.g. A330, A350) matching the same product in the model due to generation upgrade. I will still call these products B777 and A330 for simplicity when there is no ambiguity. Table 1 provides a summary of the important characteristics of the medium-sized aircraft. MD-11 is the first product in the medium-sized sector while B777 is the last to enter the market. Number of engines is an important characteristic because it is an indicator of fuel efficiency. Twin-engine aircraft are generally more efficient than aircraft with more engines.

To examine whether medium-sized aircraft can be treated as a single market, I collect route level information and calculate the following ratio for each route:

$$\text{medium-wide-ratio} = \frac{\text{total number of flights of medium wide-bodied aircraft}}{\text{total number of flights of any wide-bodied aircraft}} \tag{6}$$

If this ratio is close to 0, then it is a route where the medium-sized aircraft hardly compete with other wide-bodied aircraft (small or large); if this ratio is close to 1, then it is a route where other wide-bodied aircraft (small or large) hardly compete with medium-sized ones. However, if this ratio is close to 0.5, then medium-sized aircraft are actively competing with other wide-bodied aircraft on a given route. As such, a large proportion of routes with the ratio close to either 0 or 1 would be supporting evidence for defining medium-sized as a single market.

I observe monthly the total number of flights for each aircraft sub-model (e.g., Boeing

777-200) on any U.S. domestic and international route during the 1990-2011 period. For each month-aircraft-route observation, I also observe number of passengers, pound of freights, distance of routes, and total flying time. I focus on those routes with at least one flight of medium-sized wide-bodied aircraft and having distances longer than 1000 miles. I merge all of the post-merger years data (1997-2011) and then only keep routes that have, on average, at least 50 flights of any wide-bodied aircraft per year. All these steps are intended to help me focus on medium-sized-related routes where wide-bodied aircraft are flying in a nontrivial frequency. I also drop all non-jet observations, although they are not expected to fly on a route where wide-bodied aircraft are also flying anyway.

I end up with 908 routes. Checking the *medium-wide-ratio*, I find:

1. 61.5% of the routes with $medium-wide-ratio > 0.8$ or $medium-wide-ratio < 0.2$;

2. 74.0% of the routes with $medium-wide-ratio > 0.7$ or $medium-wide-ratio < 0.3$;

Figure 3 demonstrates the distribution of the *medium-wide-ratio*.[16] I also repeated the above steps with several single year data sets and found similar results.

I present some typical examples of routes and their major aircraft:

1. New York, NY – Shanghai, China: A340, 55.0%; B777, 45.0%

2. Miami, FL – Cologne/Dusseldorf, Germany: A330, 73.9%; A340, 11.5%; MD-11, 5.8%

3. Dallas, TX – Osaka, Japan: B777, 82.1%; MD-11, 17.9%

These markets are exclusively served by medium-sized wide-bodied aircraft. In contrast, typical routes with a medium-wide-ratio close to 0.5 are hub-to-hub domestic routes, e.g., Los Angeles to Chicago. Based on their product traits as reported in Figure 2 and demand information as reported in Figure 3, the data supports treating medium-sized wide-bodied aircraft as a well-defined market.

[16]The part of the density outside range $[0, 1]$ in the figure corresponds to observations close to 0 and 1. They are plotting bugs to be fixed. There are in fact many routes with $medium - wide - ratio = 0$ or 1. The shape of the density would remain the same after fixing the bug, except that it would be a little bit higher at 0 and 1.

2.4.2 Demand Function Estimation

Following Benkard (2004), I model yearly aircraft demand using a nested logit discrete choice model. The demand system is estimated with demand data for the period 1991-2009. A total of 12 aircraft submodels (e.g. Boeing 777-200) were observed over the period, leading to 113 submodel-year observations. Consumer a's utility function of aircraft j at time t is

$$v_{ajt} = \varphi G_{jt} + X_{jt}\beta - \alpha p_{jt} + \xi_{jt} + \zeta_{agt} + (1-\sigma)\epsilon_{ajt}, \tag{7}$$

where G_{jt} is the plane generation level measuring quality. Impact of future generations on demand is then modeled as differences in generations times φ. φ thus represents gaps in quality between generations. X_{jt} are observed characteristics including seats, maximum range, and number of engines. p_{jt} is the average price for aircraft j in year t. All prices are converted into 1994 U.S. dollars. ξ_{jt} is the unobserved component affecting demand. Its variation captures variations in consumer preference over brand and plane characteristics. Note that although characteristics are fixed for an aircraft, preference over brand or certain characteristics might change across time due to shocks such as aircraft accidents or expansion of an airline, which prefers a certain aircraft type. Since evolution of ξ is affected by these exogenous shocks, I assume generation upgrade decisions do not affect evolution of ξ. ζ_{agt} and ϵ_{ajt} are respectively the random group- and plane- specific tastes. ϵ_{ajt} is an identically and independently extreme value. I allow for two groups in the model, one includes all new medium-sized aircraft and the other includes only the outside good, which stands for small or large wide-bodied aircraft and all of the old wide-bodied aircraft on lease. $\sigma \in [0,1]$ represents the within-group correlation of utilities.

Each year is viewed as a market, and, as in Benkard (2004), the market size M is approximated by the total number of used and new wide-bodied aircraft. This approximation is consistent with the assumption that all old and new aircraft are re-sold or rented out each year.[17] If a used aircraft did not change ownership in a year, it is viewed as bought by the firm who owned it. In this sense, market size or total transaction each year equals

[17]Used aircraft trade and rental are very common. For example, almost every MD-11 airliner has changed ownership or is owned by a leasing company.

total number of used and new aircraft.

Consumer a chooses product $j \in \{0, 1, ..., J\}$ in period t if $v_{ajt} > v_{akt}$ for all $k \neq j$, $k \in \{0, 1, ..., J\}$. 0 denotes the outside good. Then integrating over the probability of choosing product j for all consumers gives the well-known formula for the market share of product j, $s_{jt} = \frac{q_{jt}}{M_t}$ as:

$$s_{jt} = \frac{e^{\frac{\varphi G_{jt} + X_{jt}\beta - \alpha p_{jt} + \xi_{jt}}{(1-\sigma)}}}{D_{gt}^{\sigma}[\sum_g D_{gt}^{(1-\sigma)}]},$$

where

$$D_{gt} \equiv \sum_{j \in \text{ group } g} e^{\frac{\varphi G_{jt} + X_{jt}\beta - \alpha p_{jt} + \xi_{jt}}{(1-\sigma)}}.$$

Taking the logarithm and rearranging terms results in the following equation to be estimated using two-stage least squares (2SLS):

$$s_{share} \equiv ln(s_{jt}) - ln(s_{0t}) = \varphi G_{jt} + X_{jt}\beta - \alpha p_{jt} + \sigma ln(s_{j/g,t}) + \xi_{jt}, \tag{8}$$

and,

$$s_{0t} = \frac{M_t - \sum_1^J q_{jt}}{M_t}.$$

Rearranging terms of Equation 8 gives the inverse demand function $P = P(Q; X, G, \xi, M)$ used in the dynamic game

$$p_{jt} = \frac{1}{\alpha}\left[\left(\varphi G_{jt} + X_{jt}\beta + \xi_{jt}\right) - (1-\sigma)\ln\left(q_{jt}\right) + \ln\left(M_t - \sum_{k=1}^{J} q_{kt}\right) - \sigma \ln\left(\sum_{k=1}^{J} q_{kt}\right)\right] \tag{9}$$

Both price and within group share $ln(s_{j/g,t})$ need to be instrumented in the demand estimation since they are likely to be correlated with unobserved aircraft characteristics ξ_{jt} (the error term in the regression). Used instruments include: observed plane characteristics, characteristics of other planes, hourly wage in manufacturing and its lagged terms, price of aluminum and its lagged terms, and number of other products within the same firm. Firm dummy variables were also tried but adding them did not improve estimation. Observed plane characteristics and characteristics of other planes are taken as instruments with the assumption that observed characteristics are uncorrelated with the unobserved components. Manufacturing wage and aluminum price are cost shifters for price and are assumed to be orthogonal to ξ_{jt}. All these instruments are widely used in the literature except for the number of other products within the same firm. Here I assume that number of other products within the same firm is not correlated with unobserved characteristics of a product. It is correlated with the price of a product because operating cost for an airline (consumer) is generally lower if its fleet consists of a set of planes from the same firm. Thus, a positive externality of products of a firm on other products in the same firm is expected. σ is identified by covariation between the within-group market share of the plane $s_{j/g,t}$ and its total market share s_{jt}. It is also instrumented by the number of other products within the same firm.

I also tried adding in other independent variables, including fuselage, first delivery year, and firm dummies, but all those variables have very small and insignificant coefficients. Besides, removing and adding them have almost no impact on estimation results.

The estimates for Equation (8) without and with the generation term G are reported respectively in Table 2 and Table 3. Note that the dependent variable s_{share} equals $log(\frac{s_{jt}}{s_{0t}})$, the percentage change of the market share ratio of product j relative to the outside good. All parameters are significant when generation G is not included in the regression. Signs of all estimates are as expected. Price has a significant negative influence on market share. Within group utility correlation σ is close to 1. Together with the comparatively larger price coefficient, it indicates high cross-product elasticities and that inside goods are closer substitutes to each other than to the outside good. The number of engines has negative effect since aircraft with fewer engines generally have higher fuel efficiency. All other factors equal (including price), airlines prefer larger planes and planes with longer range.

The last column in both tables presents standard errors of data observation as indicators of variations in the explanatory variables. Having data variation helps to determine relative importance of characteristics. For instance, in Table 2, the coefficient on maximum range ("range/10000") is estimated to be 2.04, which in absolute value is about three times as large as that of the number of engines. But the ratio of potential variation of the two characteristics is about 2/9th. Putting the coefficient and data variation together, the number of engines generally has a larger contribution to the differences in market share across products than the maximum range does. For example, take the characteristics of A330 and A340 presented in Table 1. The difference in maximum range is 0.2 ten thousand kilometers while the difference in the number of engines is 2, or ten times larger. Hence, combining the information from the first and last column, the number of engines contributes the most to market share differences among characteristics when generation G is not included. Dominance of the number of engines is understandable since it is correlated with fuel efficiency, which is a major factor in operating cost. This is also supported by the observed trend of twin-engine aircraft replacing those with three or four engines for medium-sized and small-sized wide-bodied aircraft. (With respect to the two non-twin-engine medium-sized aircraft, MD-11 was shut down after the merger and A340 experienced a low production rate in its life and ceased production in 2011.)

When generation G is taken into account, it explains most variations contributed by the characteristics, rendering them insignificant. The estimate on G suggests a 12% increase in market share ratio when generation is upgraded. As to be discussed in Section 2.4.4.1, generation differences represents differences in operating costs for airlines. Therefore, the fact that generation G has the strongest impact among characteristics (considering data variation) on market share emphasizes the important role of emphasizes the important role that airline's operating cost concern has in determining competition in aircraft manufacturing.

2.4.3 Cost Function Estimation

As with many other manufacturing industries, major variations in the unit cost of assembling an aircraft are attributable to variations in labor inputs (L). Thus, I model total variable cost (TVC) as a linear function of labor inputs L. Lockheed L-1011 is the only aircraft type that I can observe unit labor cost. I first estimate the learning curve of L-1011 and employ estimates on its total variable cost function from Benkard (2004). Benkard found the wage rate had been quite flat and fixed it at \$20/hour. Labor cost is then this wage rate times labor inputs L. Regressing total variable cost on total labor cost gives

$$TVC_{L-1011} = 36.2 + 0.12 L_{L-1011}.$$

where TVC_{L-1011} is in 1994 dollar millions and L_{L-1011} is labor inputs based on L-1011 estimates and is in 1000 man-hours.

To get the cost function of other products based on that of Lockheed L-1011, I follow the approach in the literature by assuming labor requirements per pound of aircraft is constant across planes.[18] Thus the cost function of product j can be derived from its weight ratio to L-1011, denoted as r_j. Total variable cost for product j is then calculated in the model using

$$TVC_j = 36.2 + 0.12 r_j L_j.$$ [19]

I will discuss the learning curve, L_j as a function of industry quantity vector Q and product experience level E_j, in the next section.

Fixed cost is estimated to be $200 million per year based on Lockheed's annual report on L-1011. It is a strong assumption to speculate that fixed cost is the same across products, but fixed cost has no impact on either prices or consumer surplus in a model without exit and entry. I keep fixed cost in the model only for quantifying firms' profits.

2.4.3.1 Labor Input Function

The learning curve describes the commonly observed negative relationship between accumulated production and unit labor input requirements in aircraft and many other manufacturing industries. It is decomposed into two equations in my model: labor input as a function of experience and experience as a function of current and past quantities. I will discuss the labor input function in this section and the experience accumulation function in the next one.

Following Benkard (2000), the log unit labor input requirement function for product j

[18] As Benkard (2004) pointed out, although there is no empirical evidence testing whether commercial aircraft share learning curves, literature on military production does suggest that parameters do not vary much across production lines. Further discussion on this issue can be found in Benkard (2004).

[19] I also estimated r_j using the first approach described in the beginning of this section for the model without generation upgrade. Specifically, I use difference between estimated and observed average prices as the moment condition. The estimated prices are solved from the dynamic game for each trial of r_j in searching for optimal r_j. Mimimization is carried out using KNITRO solver with its global multi-start search. I found using the weight ratio as r_j is optimal and cannot be improved.

produced at time t is estimated based on the following regression:

$$lnL_{j,t} = lnA + \gamma_1 lnE_{j,t} + \gamma_2 lnS_{j,t} + \epsilon_{j,t}. \tag{10}$$

where A is the intercept and $S = \frac{12}{7}\sum_{\tau=t-3}^{\tau=t+3} q_\tau$ is the line speed or production rate commonly included in the engineering literature.[20] As a summation of recent quantities, line speed S is endogenous and needs to be instrumented. $\gamma_2 > 1$ implies decreasing returns to scale while $\gamma_2 < 1$ implies increasing returns to scale. There is no clear implications of γ_2 without estimation since productivity of labor depends on the level of capital in the short-run. Dependence of L on experience level E highlights the learning-by-doing feature. The learning, forgetting, and spillover effect on marginal cost is then modeled as the impact of industry quantity vector Q on the evolution of experience E.

2.4.3.2 Experience Transition Function

When there is no spillover of experience across production, experience accumulation is commonly modeled as

$$E_{j,t+1} = \delta E_{j,t} + q_{j,t}. \tag{11}$$

in the literature, where learning is reflected by the positive relation between $E_{j,t+1}$ and q_t, and forgetting is modeled as $0 < \delta < 1$. Thus, experience accumulates as more aircraft are produced but also depreciates due to organizational forgetting.

I further allow a spillover effect: experience may also accumulate through production of other products. Thus, $E_{j,t+1}$ will be a function of the entire industry quantity vector Q_t. For product j, I let the contribution rate of different products on $E_{j,t+1}$ be different in two dimensions: ownership and resemblance in aircraft characteristics. The experience transition function becomes

$$E_{j,t+1} = \delta E_{j,t} + \sum_{j'}^{J} \theta_{j'} f(X_j, X_{j'}; \upsilon) q_{j',t}, {}^{21} \tag{12}$$

[20] Equation 10 can be derived from a production function with fixed capital taking the Leontief form in labor and materials. See details in Benkard (2000).

[21] The spillover effects measured by the parameters here are net effects in the sense that increases in quantities of other aircraft may also spur competition for experienced workers in the labor market. Thus,

where

$$\theta_{j'} = \begin{cases} 1 & \text{if } j = j' \text{ (i.e. on own production)} \\ \theta_1 & \text{if } j' \text{ is a different submodel of j} \\ \theta_2 & \text{if } j' \text{ is a different product in the same firm} \\ \theta_3 & \text{if } j' \text{ is a product from another firm} \end{cases} \tag{13}$$

measures the difference of across-firm spillover and within-firm spillover ($\theta_3 - \theta_2$) when products are homogeneous in characteristics. Submodels (for θ_1) are variations of a product. For example, for product A330, there are two variations, A330-200 and A330-300, which have slight differences in seats, range, and other characteristics.

$f(X_j, X_{j'}; \upsilon)$ is a product distance function. I use two characteristics: number of seats and maximum ranges.[22] Specific functional form of $f(X_j, X_{j'}; \upsilon)$ is then

$$f(X_j, X_{j'}; \upsilon) = \upsilon_1^{\frac{|X_{j1} - X_{j'1}|}{dX_1}} \upsilon_2^{\frac{|X_{j2} - X_{j'2}|}{dx_2}}, \tag{14}$$

where 1 stands for "number of seats" and 2 for "maximum range," $\upsilon_1, \upsilon_2 \in (0, 1)$; dx_1 and dx_2 are the maximum difference set to normalize the differences into $[0, 1]$. Note that from Equation (12) and (14), the larger the difference is for a given υ, the smaller the spillover effect; υ_k close to 0 implies that characteristic k has a strong impact, while υ_k close to 1 suggests that characteristic k has little impact on the spillover rate.

By substituting Equation (12) into (10), I use a GMM method to estimate all the learning curve parameters in these two equations based on monthly data of L-1011.[23] Note that $\epsilon_{j,t}$ represents the unobserved part of productivity and could be serially correlated. Since productivity interacts with choice of line speed, S, and experience accumulation, E, $\epsilon_{j,t}$ could also correlate with both E and S. Following Benkard (2000), the solution is a GMM-HAC (Heteroskedasticity and Autocorrelation Consistent) estimator suggested by Andrews (1991). The instrument variables are standard: demand shifters include various world GDP measures, the price of oil, and a time trend; cost shifters consist of the world

the parameters represents net effects of experience spillover and labor market competition.

[22]I tried fuselage and some other characteristics and the results did not change significantly.

[23]Due to the special connections between L-1011 and McDonnell Douglas's DC-10, I treat DC-10 as a within firm product for L-1011 in estimation. A detailed discussion on this choice is given in the Appendix.

aluminum price and the U.S. manufacturing wage rate.[24] Other than time trend, all shifters include both present and lagged variables.

Although an instrumental variable that specifically shifts quantity of each aircraft type is not necessary for identification, it would be helpful to have instruments that affect quantities of different types disproportionally. Thus, I include another two sets of instrumental variables. First, I use GDP growth of various regions because different regions have different demands for various aircraft types and brands.[25] Second, I use the weighted sum of all jet accidents and incidents of a firm for the previous 18 months, with less weight on narrow-bodied aircraft and freighters, divided by total aircraft in service in the same firm. See Figure 4 for an example of the negative correlation of Boeing's accidents index and quantities produced.

Parameter estimates of the learning curve are given in the first two columns of Table 4. Both characteristics have little impact on spillover (υ close to 1). Thus, I estimate another learning curve without characteristics as in Equation (15), and the result is given in the last two columns of Table 4.

$$E_{j,t+1} = \delta E_{j,t} + \sum_{j'}^{J} \theta_{j'} q_{j',t}. \tag{15}$$

Estimates are close in both cases since characteristics effects are estimated to be trivial. I drop characteristics and use Equation (15) in the dynamic part. R^2 of the estimation is 0.92. Estimated and actual labor input of each L-1011 is plotted in Figure 5. The estimates fit the data well, so I decided it is safe to ignore cost shock $\epsilon_{j,t}$ in the dynamic game. All estimates are significant except for returns to scale. The exponential of the labor cost intercept measures the unit labor requirement for the first aircraft built. As discussed before, I will make this starting level different for different models based on their weight ratios to L-1011. Thus, the shape of the learning curve is assumed to be the same while levels are permitted to be different. There is a 55% labor savings when experience

[24] For a detailed discussion on choices of these instrumental variables, see Benkard (2000).

[25] Cited in a *Wall Street Journal* article "Boeing Ups Forecast For Commercial Aircraft Demand Over 20 Years" published on June 16, 2011, Randy Tinseth, Vice President for marketing at Boeing Commercial Airplanes, said, "Economic growth, world trade and liberalization" are "the fundamental drivers of air travel" and correspondingly aircraft demand.

doubles. This seemingly large learning rate is partly offset by a high yearly forgetting rate at 43%$(= 1 - 0.9549^{12})$. Forty-three percent of experience is lost every year, making it difficult to double experience especially when experience stock is already high. This seemingly high forgetting rate is related to the relatively low aircraft production rate and customized configurations for each aircraft built. In manufacturing aircraft, assembling works repeat at a low rate and tasks are hardly ever identical. In addition, experience measures a firm's level of human capital rather than skills of each single worker. Thus, frequent turnovers due to layoffs and promotions also imply a high forgetting rate.[26] Both the high learning rate and forgetting rate imply large benefits to produce more oneself and to force one's rivals to produce less. Dynamically, there will be fierce competition among firms to reach and maintain high output and experience levels, while attempting to force others to be stuck at low output and experience.

Submodel spillover is almost complete ($\theta_1 = 0.9742$). Given this result, along with the fact that demand related characteristics are close among submodels, I decide not to differentiate submodels in the dynamic game. There is almost no cross-firm spillover ($\theta_3 = 0.0182$). This is understandable since experience is believed to be mainly accumulated through repeated practice of workers. Within-firm spillover is about a quarter ($\theta_2 = 0.2408$), indicating that building four aircraft of a different type is as helpful in experience accumulation as assembling one of the same type for a multi-product firm. Note that the large difference between within-firm and cross-firm spillover suggests potential benefits when firms merge and ownership structure changes if the within-firm spillover rate does not vary much on properties beyond ownership.[27]

[26] This forgetting rate is close to the 39% forgetting rate estimated in Benkard (2000). See Benkard for further discussion on the high forgetting rate.

[27] Several circumstances contribute to a large within-firm spillover effect. First, internal shifts of the workforce are much easier than shifts across firms, and a firm may reallocate workers among different departments to improve efficiency. Second, internal shifts help firms to avoid organizational forgetting by keeping workforce busy assembling other models when demand for a certain model is temporarily low. Furthermore, managerial ability and labor-cost-related production techniques usually can be shared only within a firm, either due to firm differences or the need to keep business secrets.

2.4.3.3 Discretization of Experience

With estimates given in Table 4, next period experience can be calculated using Equation (15) for given experience and quantities of all products in a period. Experience defined in Equation (15) is a continuous variable. To apply it as a state variable in the dynamic game, I discretize the experience variable for each product into 7 grids:

$$\mathbb{E} = \{1, 10, 20, 40, 70, 110, 165\}.$$

I use E^k to denote experience at the k*th* grid. (e.g., $E^2 = 10$.) With enough grids, the experience process can be approximated arbitrarily well. I will explore that more in Section 2.4.5.2.

I denote the experience level resulting from Equation (15) as $E^*_{j,t+1}$, namely,

$$E^*_{j,t+1} = \delta E_{jt} + \sum_{j'}^{J} \theta_{j'} q_{j'}$$

Then the experience transition process is modeled as

$$E_{j,t+1} = \begin{cases} E^u_j & \frac{E^*_{j,t+1} - E^d_j}{E^u_j - E^d_j} \\ E^d_j & 1 - \frac{E^*_{j,t+1} - E^d_j}{E^u_j - E^d_j} \end{cases} \tag{16}$$

E^u_j is the smallest grid in $\mathbb{E}$ larger than $E^*_{j,t+1}$, and E^d_j is the largest grid smaller than $E^*_{j,t+1}$. Thus, both E^u_j and E^d_j also depend on quantity Q_t.

2.4.4 Generation Upgrade

In this subsection, I first present how generation is defined in the wide-bodied aircraft industry. Then I discuss employment of relative vs. absolute generation for the medium-sized wide-bodied aircraft industry. Calibration of related parameters are presented at the end.

2.4.4.1 Generation and Generation Upgrade

Ideally, I would treat each new aircraft type as a new product and allow a specific new vector of characteristics for it. However, it is impossible to do so as we do not observe characteristics of products not yet introduced. Instead, I categorize aircraft by generations according to some criteria. Average generation gap φ is estimated in Section 2.4.2. Note that for the purpose of quantifying the merger effect on expected future welfare and the upgrade rate, knowing the average generation gap is sufficient.

There are many ways to define a generation of a jet airliner. Loosely speaking, a new generation has substantial demand-side advantages attributed to more desirable characteristics over the old generation. Empirically, one simple way is to treat each aircraft submodel (e.g. the Airbus 330-200) as a new generation. However, differences between some types of aircraft are quite distinctive from the differences between other types. New aircraft type and submodels have been introduced to provide longer range, different options in size, higher fuel efficiency, lower CO_2 emission, etc. These variations create discrepancies in defining a new generation as any new model introduced. In addition, the demand effect of some new models are small and defining every small changes as a new generation will result in too many possible relative generation levels that again causes the dimensionality problem. Definition of generation needs to be applicable as well as capturing major demand effects. The wide-bodied aircraft industry had evolved for 28 years before the merger and there was no room left for firms to introduce new aircraft with range or plane size meeting the market demand that is not yet covered by an existing type. Models introduced after the 1997 merger were generally driven by concern over operating cost.[28] Hence, I define a new generation of aircraft family as one that provides at least 5% lower operating cost for airlines. In the medium-sized wide-bodied aircraft market, introduction of Boeing 777, 787, and Airbus A350 are treated as new generations according to this definition.[29]

Upgrade in generation may also involve a huge amount of one-time development cost and the lowering of experience level due to adjustments in production procedures. Thus,

[28] This is confirmed by discussions with Edmund S. Greenslet, an aircraft industry expert, and publisher of *The Airline Monitor*.

[29] B787 and A350 are new generations of B777 and A330, respectively.

generation upgrade provides a second source of experience setback other than organizational forgetting. As discussed above, since a product will have different characteristics when its generation is upgraded, its experience level will be lower as workers and technicians will not be familiar with the new specification; it takes practices to figure out new mechanisms that suit the new generation. One such evidence is the temporary rise in labor requirement for the introduction of Lockheed's L-1011-500.[30] Figure 6 plots labor input (in 1000 man hours) for each L-1011 built. There were only minor changes in characteristics involved in the introduction of the L-1011-500, which does not even qualified as a new generation according to our definition. However, we can still see a clear cost rise for the first two L-1011-500 aircraft produced in the figure. Following the initial rise, the -500 type required a slightly higher labor input and the difference vanished eventually. Vanished difference suggests that the initial cost rise for the -500 type is not due to a systematically higher cost requirement but resulting from a temporary lowering of experience due to the introduction of new types. The detrimental effect of innovation on experience accumulation is also supported by the work of Levitt, List, and Syverson (2012) in the automobile industry as discussed in the introduction. I will discuss the calibration of upgrading cost and determination of experience setback later in this section.

2.4.4.2 Relative vs. Absolute Generation

A modeling obstacle of introducing generation upgrade into a dynamic game is that generation itself needs to be a state variable for each product, and it is implausible to assume that the highest generation exists. In fact, even if I assume there is a highest generation level, there are at least two problems. First, existence of the highest level implies that the dynamic mechanism in generation is lost when all products reach the highest level. Second, it is impossible to limit possible generations to a small number, which leads to a dimensionality problem.

I choose to deal with the dimensionality problem by assuming that only relative quality (or generation) matters in determining the individual demand function for each product. Therefore, I can use the pace of evolution of the outside good as a benchmark, and track

[30] Benkard (2000) also published the same finding.

only the generation difference of each product relative to the outside good.[31] More precisely, the benchmark is the average difference between the inside goods and the outside good in generation. Thus, generation state G in the model is the difference from that benchmark average difference. Because I observe no more than 2 generation lags among all inside goods (new medium-sized wide-bodied aircraft), I model the generation difference state variable $G_j \in \{-1, 0, 1\}$, where -1 denotes one generation behind the average difference between inside and outside goods and 1 denotes one generation ahead of it. In this sense, fixing product and product characteristics is equivalent to assuming that the inside goods and the outside good have the same pace of technology improvement. In modeling, this is to fix generation difference state variables for all products at 0.

Theoretically, the outside good is a composition of any products that can be viewed by some consumers as substitutes of the products in the market. Therefore, with respect to the medium-sized wide-bodied aircraft industry, the outside good could consist of old wide-bodied aircraft for sale and new non-medium-sized wide-bodied aircraft and even narrow-bodied aircraft. Different components have different importance in terms of their degree of substitutions to the inside goods.

Since production decisions on non-medium-sized wide-bodied aircraft are also made by Airbus and Boeing, the evolution of the outside good is partially endogenous. However, the event of relative generation downgrade of all inside goods has broader interpretations than the generation upgrade of the outside good. Examining Equation (9) infers that downgrade of G_j for all inside goods is equivalent to any permanent negative shock on overall demand that lowers prices for all products by $\frac{\varphi}{\alpha}$. For the evolution of demand and generations in an industry, by simply observing the evolution itself, it cannot be determined whether it is driven by innovations in the outside good or by a permanent negative demand shock. In fact, there is no need to make such distinctions in determining generation evolution of the inside goods. For the medium-sized wide-bodied aircraft industry, generation upgrade is mainly driven by macro economic shocks (e.g., the desire for better fuel efficiency due to rising petroleum prices after the September 11th attack), increasing demand for international

[31] The idea of relative generation is originated from relative quality modeled in Pakes and McGuire (1994) and is similar to that in Goettler and Gordon (2011). See further discussion in the introduction.

travel qualities, and increased supply in related markets. For example, Chinese manufacturers recently entered the narrow-bodied aircraft industry. This event is equivalent to the outside good moving to a new generation in the sense that both will permanently shift demand away from the medium-sized wide-bodied aircraft industry and stimulate Boeing and Airbus to innovate for more attractive planes. All these mechanisms are exogenous to the medium-sized wide-bodied aircraft industry; hence, evolution of the market-wide generation downgrade can be treated as exogenous.

2.4.4.3 Generation Related Parameters

I first specify distribution of upgrading cost c_j^G and experience setback function $\psi(E)$ for applying the model to the aircraft industry. c_j^G is assumed to be drawn from a uniform distribution $U[C^d, C^u]$.[32] The largest and smallest development cost of recent new aircraft models and submodels are chosen as $C^d = 330$ and $C^u = 614$ (in 1994 dollar millions). The experience setback function $\psi(E)$ is more difficult as estimating it in the first stage requires observation of a products's unit labor requirement and generation upgrade choices. Experience setback in L-1011-500 provides a lower bound of setback magnitude, but I do not have sufficient data to pin down a specific value. The strategy is to assume generation upgrades setback experience by n_G grids for the discrete experience state introduced in Section 2.4.3.3. That is,

$$E_{j,t} = \psi(E_{j,t-1}) = \min\{E^{g-n_G}, E^1\}, \tag{17}$$

where E^g is the discretized grid that E_j lands on with $g \in \{1, 2, ..., 7\}$. Varying n_G would demonstrate impact of setback magnitude on firm behaviors. In the merger evaluation, because there are not enough instances of generation upgrade observed to fully estimate the upgrade cost, I set $n_G = 1$.[33]

Finally, p^G, which also represents the industry long-run innovation rate, is obtained as

[32] I can also model C^G as a choice variable rather than a random draw. Randomness is then introduced through probability of success generation advance, which increases with upgrading cost. Although this alterative is more common in the literature, it is less attractive for the aircraft industry since larger investment in aircraft development is realized over time and is generally related to unexpected difficulties in development rather than higher probability of success.

[33] I also tried setting n_G to 0 or 2 and found no significant differences in quantifying merger efficiencies.

the inverse of the average years across product before generation advances, which is 10.75. Thus, $p^G = 1/10.75 = 0.09$.

2.4.5 Dynamic Game Specifics

In this sub-section, I explore three issues related to the dynamic game. First, the preference rank state variable is examined in its role to reduce the state space and lessening the computational burden. It can be employed in a dynamic game of any industry with many products that can be grouped into limited categories based on unobserved characteristics. Second, I provide a test with respect to concerns on sensitivity of choices of discretization of the state variables. I conclude this part with arguments on why exit and entry need not to be directly modeled for the medium-sized wide-bodied aircraft industry. Those readers who are not interested in these issues may want to skip to Section 5.

2.4.5.1 State of Preference Rank

Demand estimation provides a panel data of unobserved characteristics ξ of all products. Fluctuation in ξ represents changes in consumer taste driven by exogenous fluctuation in various sources, such as important accidents or technological problems specific to a product or a firm, personnel changes in important airlines, operating-cost-related macroeconomic shocks that lead to preference of twin-engine aircraft, and the temporary spur in international travel driven by the business cycle that makes relatively larger planes more attractive, etc. Ideally, I would model this exogenous fluctuation in ξ by allowing ξ_j to be a state variable for each product and then estimate its stochastic process using cell means. However, adding in one more state variable for each product results in the well-known "curse of dimensionality" problem. Figure 7 provides a histogram of percentile prediction error for the market share ratio s_{share}. Predictions errors are smaller than 3 percent for most observations and smaller than 10 percent for all. This suggests that the error term ξ is marginal in explaining variations in market share in demand estimation.[34] So to save computational power for more important aspects, I compromise by putting restrictions on joint transitions

[34] The impact of ξ on quantity prediction is relatively large, but the majority of the prediction error is still less than 15% as in Figure 8.

of all ξ_j and introduce the preference rank state variable discussed below.

For the model of the medium-sized wide-bodied aircraft industry, macroeconomic shocks influencing the entire market are captured by market size state variable M; change in product qualities are captured by the generation upgrade decision; observed differences in characteristics are captured by X. Suitability is not an important issue when working with the medium-sized market instead of the entire wide-bodied market. Then, variation in ξ is most likely driven by two other factors: variation in preference over the more fuel efficient twin-engine types and variation in preference over firm brands. Therefore, I assume that variation in ξ_j, denoted as $\Delta\xi_j$, can be decomposed into two additive parts that evolve independently, with

$$\xi_j = \overline{\xi}_j + w_j^T \cdot \kappa_j^T + w_j^F \cdot \kappa_j^F, \tag{18}$$

where

- $\overline{\xi}_j$ is mean value of time series ξ_j;
- w_j^F and w_j^T are given weights;
- κ_j^T is variation of preference over twin-engine types; and
- κ_j^F is variation of preference between Boeing and Airbus products.

κ_j^T and κ_j^F are preference rank state variables used in the dynamic game. κ_j^T is common among twin-engine types and κ_j^F takes the same value for products of the same firm. I denote the vectors of κ_j^T and κ_j^F for all types and all firms as κ^T and κ^F, respectively. Thus, the lengths of both vectors depend on the number of firms/types instead of the number of products. For example, κ^T is of length 2 because I have two types: “twin-engines” and “not-twin-engines.” The two *vectors* κ^T and κ^F are *preference rank* state variables in the dynamic game and evolve stochastically over time.[35] In the dynamic game, I allow each vector to take on two possible values. Specifically, $\kappa^T = \kappa^{T1}$ is the vector for state where

[35] Note that transition of the preference rank state variable is over the entire vector κ^T instead of each element of it. This further reduces the number of state variables. The idea behind this is that demand is determined by relative ξ_j among products. Thus, relative preference over firms and types captures major information of its absolute value counterpart. Finally, note that variation of preference of product j relative to the outside good is captured by the variation of market size M.

twin-engine aircraft are relatively preferred while $\kappa^T = \kappa^{T0}$ is the vector when they are less attractive. Similarly, $\kappa^F = \kappa^{F1}$ when Airbus is preferred while $\kappa^F = \kappa^{F0}$ when Boeing is preferred.

Using the panel data of unobserved characteristics ξ, the parameters $(\bar{\xi}_j, \kappa_j^T, \kappa_j^F, w_j^T, w_j^F)$ and transitions of κ^T and κ^F are calibrated as follows:

1. $\bar{\xi}_j$ is calibrated as the mean of time series ξ_{jt} and variation in ξ is calculated as

$$\Delta\xi_{jt} = \xi_{jt} - \bar{\xi}_j.$$

2. Among the four products in the model, A330 and A340 are of the same firm but only A330 is a twin-engine. Based on Equation (18), differences in series of $\Delta\xi_{A330}$ and $\Delta\xi_{A340}$ come from engine-difference only. κ^T and its transition are calibrated as
   (a) For each time t, if $\Delta\xi_{A330,t} \geq \Delta\xi_{A340,t}$, $\kappa_t^T = \kappa^{T1}$. Otherwise, $\kappa_t^T = \kappa^{T0}$.
   (b) Using the time series of κ_t^T from (a), its transition matrix is then estimated in the usual way of a Markov chain.
   (c) Value of κ^{T1} is chosen as the conditional mean $((\overline{\Delta\xi}_{A330}, \overline{\Delta\xi}_{A330})|\Delta\xi_{A330,t} \geq \Delta\xi_{A340,t})$.[36] The same applies to κ^{T0}.

3. κ^F and its transition are calibrated similarly using the time series of $\Delta\xi_{A330}$ and $\Delta\xi_{B777}$, both of which have two engines.[37]

4. w_j^T and w_j^F are chosen to minimize the distance between observed and calibrated panel of $\Delta\xi$.

Figure 9 demonstrates the fit of data for ξ calibrated from Equation 18 (labeled as "Rank") and for having a binary state variable for each ξ_j (labeled as "Cell"). The preference rank approach is able to provide better predictions for transitions and no worse fitting in values, while reducing the size of the state space. Parameters and transition matrices estimated are given in Table 5.

[36] If there is more than one product of the same type, quantity weighted means can be used instead.
[37] Value of κ_{MD11}^F is also conditional on whether Airbus or Boeing is preferred. I tried to allow preferences rank states over three firms, but did not find much difference.

Recall that ξ represents airline preference over brand and characteristics. Although it is reasonable to assume that a merger does not affect airline preference over characteristics, it certainly changes product ownership. This creates problems on how one should adjust ξ_j if it were modeled as a state variable with its transition estimated based on its own time series. However, with the introduction of the preference rank state variable, the merger's impact on ξ_j through ownership change is directly captured by κ_j^F in Equation (18).

2.4.5.2 Sensitivity of Discretization

Both experience state variable E and market size state variable M are discretized. Solving quantity choices at all state profiles can be viewed as a non-parametric approximation of the underlying equilibrium function from state space to policy space. Then a natural question is whether I have chosen enough number of grids so that the approximation is close to the underlying function. Although it is impossible to test sensibility of the choice of the number of grids by having infinitely many grids, robustness can surely be tested by, for example, doubling the number of grids and comparing resulting policy functions. I tried this on several model specifications and with various denser griding methods and found close equilibrium policy functions. Hence, it is reasonable to believe that the result is robust to the discretization method. While state space has high dimensions, demonstrating policy function for more than three dimensions in one figure is hardly instructive. Thus, I chose to plot one policy variable on two varying state variables while keeping other states fixed. This led to hundreds of figures to cover the entire policy function even for very basic model specifications.

Representative of that output, Figure 10 plots quantity of Airbus A330 as a function of experience levels of Airbus A330 and A340, fixing experience level of Boeing 777 at the lowest grid and market size at the highest grid.[38] The blue plane is solved from the model employed in the paper where E is discretized into 7 grids while the red plane is solved from a model with a denser grid for E by adding a grid point between any of the seven original

[38] The equilibrium strategy for Figure 10 is solved from a scenario of post-merger with the MD-11 shut down for the model without the generation upgrade feature. Figures for firm strategies with respect to all states are available upon request.

grids for E, i.e.,

$$\mathbb{E}' = \{1, 4.5, 10, 15, 20, 30, 40, 55, 70, 90, 110, 137.5, 165\}.$$

The two planes are generally close to each other, suggesting the choice of seven grids provides a close approximation to the underlying policy function.

2.4.5.3 Exit and Entry in the Medium-sized Wide-bodied Aircraft Industry

Entry and exit decisions of both firm and product levels are generally assumed away in the model. The only exception is that I allow a firm to switch any of its product to a potential entrant good by setting quantity of that product to 0[39] in any period and reverse the process by setting a positive quantity in any future period. Here I elaborate on reasons why there is no need to directly model entry and exit.

There are at least two reasons why entry on the firm level is rare in the wide-bodied aircraft industry. First, it requires huge initial capital and a complete set of frontier technologies to start a new business. Historically, all four firms that participated in the wide-bodied aircraft sector have been active in other sectors of the aircraft industry and are somehow subsidized by powerful governments in their military sectors. Second, the state of the art technologies employed in aircraft design and manufacturing also work as entry barriers. Third, the learning curve feature acts as an entry barrier since it implies that an entrant cannot make any profit until after a long period. Moreover, a firm may incur a potential huge loss if there are not sufficient sales later at the bottom of the learning curve to reimburse pricing below marginal cost in the early stage. Business failure of Lockheed L-1011 stands as a perfect example and a live lesson. It is also the only incidence of exit not resulting from a merger in the industry.

No new firms have entered the wide-bodied market since shortly after the industry spawned about 40 years ago. Given that developmental time needed is at least five years and no entering intention has been revealed by 2012, it is safe to say that there will be no

[39]In the empirical model with the demand function defined based on discrete choices, optimal quantity choices are never 0, but can be arbitrarily close to 0. In this case, I call the quantity "effectively" zero since it only has negligible difference from an absolute 0.

new entrant until at least by 2017, that is, 20 years after the merger studied here. As for exit, for the only two remaining firms, Airbus and Boeing, no evidence exists that they will exit the market, particularly considering their important political strategic status.

Although entry and exit on the product level is not directly modeled as a firm choice variable, I allow it in restricted format. First, a product can be switched between a potential entrant and an active good through quantity choices as discussed above. Note that in the model, when quantity of a product is effectively zero, it has no effect on choices of other products or consumer surplus. This is demonstrated by the empirical results on MD-11 shortly after the merger in the scenario where MD-11 is not immediately shut down after the merger. Second, the model is also consistent with the introduction of new aircraft as future generations of the current types. This is because products in the model are captured by characteristics, and their advances are captured by generation upgrade. So introduction of a new generation model replacing the old one is viewed as a quality improvement of a product. Hence, I feel comfortable to assume away exit and entry and instead focus on experience and generation evolution that I believe are much more important in the dynamics of the industry.

2.5 Results of Dynamic Analysis

Now that the dynamic model for the medium-size wide-bodied aircraft industry has been setup, it can be used to address questions with regard to the impact of the 1997 Boeing-McDonnell Douglas merger. First, I ask questions about the merger effect on consumer welfare. What is the net effect of the merger on consumer surplus? How much efficiency comes from dynamic mechanisms of learning-by-doing and generation upgrade? Was the merger efficiency primarily attributable to learning-by-doing or generation upgrade? If the dynamic mechanisms were ignored and instead a traditional static model were used, how would the answers differ? Second, I examine how the merger affects firm behaviors and market structure. How did the merger affect experience accumulation and generation upgrade? If we had forbid Boeing to shut down the MD-11 immediately after the merger, would Boeing have found it profitable to keep the MD-11 in the long-run? What would have been the impact of keeping the MD-11? I address all of these questions in Section 2.5.1.

Third, recent innovation events suggest that the aircraft upgrade rate and magnitude are likely to be systematically higher in the future. I thus perform comparative static analysis to examine the impact of this possibility. Specifically, what would the net consumer welfare for the merger be if the generation upgrade rate and magnitude were larger? Section 2.5.2 deals with this question.

To address these questions, I solve three types of games:

- *Game A:* Dynamic game with learning-by-doing and generation upgrade
- *Game B:* Dynamic game only with learning-by-doing
- *Game C:* Static game without learning-by-doing or generation upgrade[40]

Game A corresponds to the full model described in Section 2.2 while Game B and C remove features from the full model to isolate learning-by-doing and market power effects. All dynamic effects are assumed away in Game C, so the merger effect in it reflects only the market power effect. The difference between Game B and C then reveals the impact of dynamic learning-by-doing. The influence of generation upgrade can be studied by comparing results from Game A and B. Finally, Game C is a traditional static model, so comparing it with Game A also reveals potential bias when ignoring dynamic mechanisms.

To evaluate merger efficiency, I solve dynamic models for three different industry scenarios for each of the three games:[41]

- *Scenario (i):* Boeing merged with McDonnell Douglas and immediately shut down MD-11 (which is what actually occurred)
- *Scenario (ii):* Boeing kept MD-11 after the merger.[42]
- *Scenario (iii):* No merger

The effect of the merger is quantified by comparing Scenario (i) and (iii). The comparison of Scenario (i) and (ii) examines the difference between forcing MD-11 to be shut down

[40] For calculation of total discounted values, it is assumed that the same static game repeats in every period.

[41] Discount factor ρ is set to 0.925.

[42] In this scenario, Boeing can set quantity of the MD-11 to 0 and pay the fixed cost. MD-11 then will function as a potential entrant for Boeing that can come back in production at any time.

immediately after the merger and letting it evolve endogenously after the merger. For each scenario, with the solved equilibrium strategies, I compute the time series of expected values of price, quantity, experience stock, upgrading probability, profit, consumer surplus, and total surplus for 50 years starting from the state of 1997.[43]

Because Hicksian and Marshallian demand functions are identical in the nested logit discrete choice model, consumer surplus can be obtained simply by integrating the demand function.[44] Following the literature (See Small and Rosen (1981) or Trajtenberg (1989).), the formula for consumer surplus is:

$$CS = \frac{M \cdot ln(1 + (\sum_j e^{\frac{\varphi G_j + X_j \beta - \alpha p_j + \xi_j}{1-\sigma}})^{1-\sigma})}{\alpha}.$$

Note that the CS formula above does not account for consumer benefits from absolute generation upgrades. This is not a problem for the merger evaluation as those benefits will cancel out when comparing the merger scenario with the no-merger scenario. In addition, consumers' preferences on product qualities also evolve over time. Thus, CS can be viewed as consumer surplus adjusted for demand evolution, with a rate assumed to be the same as the industrial innovation rate p^G.

One-time cost synergy of the merger is modeled as an experience stock transfer from MD-11 to Boeing 777 with a transfer rate τ. The merger is also likely to have fixed cost synergies, although these cannot be estimated with only one observed merger in the aircraft industry. However, fixed cost synergies do not affect price or consumer surplus. Formally, experience transfer follows the equations:

$$E_{B777}^{Post-Merger} = E_{B777}^{Pre-Merger} + \tau E_{MD11}^{Pre-Merger};$$
$$E_{MD11}^{Post-Merger} = E_{MD11}^{Pre-Merger} + \tau E_{B777}^{Pre-Merger}.$$

With $\tau = 0$, no experience stock is transferred; with $\tau = 1$ all experience stocks are transferred. Under the scenario where MD-11 is kept after the merger, the above equations

[43] All paths for different scenarios converge within 25 years. The years after reaching convergence have no impact on comparisons across scenarios.

[44] Change in CS is a compensation variation, and CS tends to be overestimated in a logit-based model.

assume that experience stocks are symmetrically transferred between MD-11 and Boeing 777. Asymmetric transfer rates can be easily incorporated in the model if necessary.

Since experience transfer is essentially a one-time experience spillover across products when product ownership changes, a potential benchmark for τ is the estimated difference between the within-firm spillover rate and the cross-firm spillover rate, i.e., $\theta_2 - \theta_3$. However, there is not enough evidence to conclude such a relationship since the underlying mechanisms of one-time sharing might be different from experience sharing each period. For example, building an aircraft involves thousands of tasks and different firms might excel in different tasks. A merger then helps sharing advantages in different tasks that might give a larger boost in cost reduction than common experience spillover by having workers perform similar tasks on different planes. I then choose to compute the welfare effect for all $\tau \in \{0, 0.01, 0.02, ..., 1\}$ to evaluate the effect of the experience transfer rate.

In the rest of this section, I discuss the impact of the 1997 Boeing-McDonnell Douglas merger on consumer welfare and market structure (Section 2.5.1). In light of the recent spur in innovation for low operating cost aircraft in the 2010s, I then perform a comparative static analysis on the impact of innovation rate and magnitude on merger efficiency (Section 2.5.2). The comparative static also serves as a sensitivity check for quantified merger efficiency with respect to generation upgrade parameters.

2.5.1 Merger Evaluation

Tables 6-8[45] present total discounted surpluses and profits for the three scenarios for game A, B, and C, respectively. For games A and B that have learning-by-doing, the merger scenarios are further categorized into two cases: no experience transfer ($\tau = 0$) and complete experience transfer ($\tau = 1$).

Larger experience transfer helps Boeing to lower its marginal cost. Thus, when τ increases from 0 to 1, both Boeing and the consumers should be better off. However, when Boeing has a large cost advantage under $\tau = 1$, Airbus products' might making less profits.

[45]Negative profits arise in the tables for products that are effectively not active in production. The discrete choice model of the demand function leads to a production level close to, rather than equal to, 0, when a product should be shut down. Since I do not allow exit, the resulting profit is close to total discounted fixed cost. It does not affect estimations of prices or consumer surplus.

These intuitions are confirmed comparing columns for $\tau = 0$ with those for $\tau = 1$ in Tables 6 and 7. When there are dynamic efficiencies (Tables 6 and 7), the merger lowers consumer surplus when $\tau = 0$ but raises it when $\tau = 1$. When there is no dynamic efficiency, Table 8 suggests that the merger is detrimental to consumer welfare. Finally, comparing results between Scenario (i) and (ii) in all three tables implies that consumers are always better off when MD-11 is kept after the merger. However, keeping MD-11 lowers Boeing's total profit.

Analyzing Tables 6-8 leads to the following property with respect to the merger.

Property 1 (The Welfare Effect Property). *With complete transfer of experience, the merger increases consumer surplus by $1.57 billion while the static equilibrium model predicts a $22.53 billion loss. Merger efficiency mainly comes from learning-by-doing. The presence of generation upgrade raises net merger efficiency at a smaller τ but reduces it at a larger τ. The merger has no impact on long-run consumer welfare.*

The last 2 columns of Table 6 indicate that the merger effect on consumer surplus is $1.57 billion when experience transfer is complete and $-1.78 billion when there is no experience transfer. The entire relationship between the net consumer surplus and experience transfer rate τ is demonstrated in Figure 11. Values of surpluses, prices, and profits for this and all subsequent figures are in millions of 1994 dollars. The horizontal line in the figure marks zero consumer effect for the merger. The solid and dashed curves plots net consumer surplus for Game A and B, respectively. The solid curve for Game A increase with τ and crosses the horizontal line at around $\tau = 0.2$, inferring that the merger is beneficial to consumers as long as there is at least a 20% transfer rate. Note that the break-even spillover rate $\tau = 0.2$ is smaller than the difference between the within-firm spillover and the cross-firm spillover ($\theta_2 - \theta_3 \approx 0.22$, see Table 4.).

The last column of Table 8 shows that abstracting away dynamic efficiencies, the pure market power effect leads to a consumer loss as large as $22.53 billion. Since net consumer surplus in the full model (Game A) is the difference between merger efficiency and the market power effect, a large market power effect implies that absolute efficiency from learning-by-doing and generation upgrade is also large. In addition, Game C corresponds to the traditional static analysis. The large difference in consumer surplus between Table 6 and Table 8 suggests that ignoring dynamic effects can lead to seriously biased results and erroneous conclusions with regard to the welfare impact of the merger.

Comparing the two curves in Figure 11 and the last two columns in Tables 6 and 7 suggests that the presence of generation upgrade raises net merger efficiency at a smaller τ but reduces it at larger τ. On the one hand, by shutting down MD-11, productions are concentrated on other products with higher quality after the merger. This channel of merger efficiency is not captured for the model without generation upgrade (Game B). From this prospective, adding generation upgrade to the model increases consumer surplus after the merger. On the other hand, when experience is higher, quantity is also higher, inferring a larger loss for a given unit cost rise. Since experience is higher with the merger, loss from experience setback due to generation upgrade is then larger with the merger. In this sense, generation upgrade erodes merger benefits from learning-by-doing. Modeling generation upgrade thus leads to two opposite forces on merger efficiency. When there is no experience transfer, experience is low and the setback effect on it is minimum. Thus, the effect of concentration on the higher quality product dominates, and net consumer surplus is larger for the model with generation upgrade (Game A) at $\tau = 0$. As experience transfer rate τ increases, the experience setback effect becomes more and more important. At $\tau = 1$, the results suggest that experience setback effect dominates and the presence of generation upgrade lowers merger efficiency. Finally, differences in net consumer surplus between Game B and Game C (about \$20 billion) is much larger than that between Game A and Game B (about \$2 billion). Thus, the primary efficiency comes from the difference between Game B and Game C, that is, the effect of learning-by-doing rather than generation upgrade.

Figure 12 reports the evolution of expected consumer welfare since 1997 for each scenario. The merger has only an intermediate influence on consumer welfare; per period consumer surpluses are the same in the long-run for all scenarios. This is because MD-11 is not in production and market structures converge to the same steady state for all scenarios in the long-run. However, two things needs to be clarified regarding absence of the long-run effect. First, it does not render the dynamic analysis futile. In the absence of the intermediate dynamic efficiency, a static model leads to misleading conclusions on consumer welfare.

Second, the long-run effect result here is specific to the Boeing-McDonnell Douglas merger, and is particularly due to the inferiority of MD-11. In general, long-run effect is likely to emerge for a different merger in a dynamic analysis. I will discuss more on this in the next property.

Property 2 (**The Market Structure Effect Property**). *Only A330 and B777 are actively in production in the long-run in all scenarios. For the first several years after the merger, the merger accelerates experience accumulation but has no clear implication on the innovation rate. The merger has no impact on long-run firm behavior.*

Figures 13-16 and 17-20 report the evolution of expected quantity and experience respectively since 1997 for each scenario for the four products A330, A340, B777, and MD-11. Recall that experience is simply accumulated quantities through the mechanism of learning, forgetting and spillover. The first observation that can be made from the four quantity figures is that only A330 and B777 are in production in the long-run.[46] The results suggest A340 and MD-11 are less favored by airlines than A330 and B777 are, and their market shares were gradually eroded by competitors. Learning-by-doing reinforces disadvantages of A340 and MD-11. On the one hand, low production rates of A340 and MD-11 are not enough to cover organizational forgetting in experience, leading to rising marginal cost. On the other hand, the competitive products, A330 and B777, are able to achieve lower marginal cost through learning because of the added production due to the merger. Enlarged differences in marginal cost eventually drive out A340 and MD-11. This result is consistent with the reality where MD-11 was shut down immediately after the merger and A340 phased out in 2011. Exit of A340 comes early in the model because the model prediction provides expected path while the reality path is just one realization possibly affected by several positive shocks.[47] The mechanism of learning and forgetting favors concentrated production from the prospect of reaching and maintaining high experience levels. If product distances are not large enough, it is not profitable to keep two similar products in one firm.

Now turn back to the figures reporting experience. For $\tau = 0$, experience levels of A330,

[46] Experience stock of A340 and MD-11 (when merged into Boeing) are not 0 because of experience spillover defined in Equation (15). They have no impact on market evolution as long as they are not in production.

[47] Recall the discussion in the beginning of Section 2.4 on the time series being one observation in a dynamic model.

A340, and B777 are slightly higher for a few years if the merger took place. However, the merger benefit of experience accumulation is small since production of MD-11 is low and quickly approaches 0 even without the merger. (See Figure 16.) This explains why learning-by-doing is not large enough to cover the market power effect if $\tau = 0$. However, for $\tau = 1$, B777 would enjoy an intermediate marginal cost advantage if the merger took place. Cost advantage of B777 would be so large that it would lower quantities and experience levels of Airbus products. In the long-run, MD-11 would not be in production whether there was a merger or not. In addition, Boeing's cost advantage is not large enough to discourage Airbus from catching up. Thus, all scenarios, with or without merger, converge to the same steady state in the long-run.

Figures 21-24 demonstrates path of upgrade probabilities. A340 and MD-11 are not produced in the long-run, so there is no upgrade on them. Long-run upgrade rates for A330 and B777 are equal to the outside good upgrade rate as inherited in the model; firms only upgrade to maintain optimal generation levels in the long-run. Figure 23 indicates that for the first 5-6 years after the merger, upgrade of B777 is more likely to take place earlier for the merger scenario with complete experience transfer than for the no-merger scenario. This is probably because Boeing would have a cost advantage high enough right after the merger that a little setback in experience in exchange for a higher quality is profitable. Generally speaking, the effect of the merger on generation upgrade is ambiguous. After the merger, softened competition could discourage innovation but enlarged market share may mean a bigger benefit from a better quality product, which would stimulate incurring the fixed cost to innovate. In addition, generation upgrade negatively impacts experience and raises unit cost, which further complicates the impact of the merger on upgrade decisions.

Property 3 (The MD-11 Property). *With the merger, consumers are worse off with immediate shutdown of MD-11 (Scenario (i)) compared to continuing production of MD-11 (Scenario (ii)). However, total profits of the merged firm would be lower and would need to be subsidized to keep MD-11. In addition, with the merger, MD-11 is phased out faster.*

Recall that the learning effect is potentially beneficial for the merger, either through concentrating learning on fewer products and reaching the bottom of learning curve faster in the scenario where MD-11 was shut down immediately after the merger, or through within-firm spillover when MD-11 is kept after the merger. Thus, keeping MD-11 after the merger might benefit consumers by enjoying experience spillover while avoiding reduced number of products. Comparing consumer surplus and product profits for Scenarios (i) and (ii) in Tables 6 and 7 shows that keeping MD-11 leads to a higher consumer surplus but lower profits for Boeing and Airbus. Theoretically, keeping MD-11 might also be beneficial for Boeing if there were sufficiently large spillover effects and significant differences in characteristics between MD-11 and Boeing 777. However, the results suggests that Boeing's total profit would be lower because it would incur fixed costs from MD-11 that could not be fully covered by revenues from MD-11. Airbus's profit would also be lower because it would face more competition. Thus, if a policy maker wanted to keep MD-11 for consumers' benefit, Boeing would need to be subsidized.

In Figure 16, quantity curves for the merger scenarios are lower than the curve for the no-merger scenario for the first 6-7 years after the merger. Namely, MD-11 would phase out faster if it was merged into Boeing. On the one hand, MD-11 receives more experience spillover after the merger. On the other hand, Boeing needs to internalize business stealing of the more promising B777 from production of MD-11. The result shown in Figure 16 indicates that business stealing concerns dominated and Boeing found it more profitable to concentrate on production of Boeing 777.

2.5.2 Comparative Statics

The 2010s is witnessing a boom of generation upgrades in the entire aircraft industry. For the medium-sized wide-bodied market, Boeing 787 was introduced to replace 777 in 2011 and Airbus responded with the new A350, which is projected to take over A330's market in 2014. Boeing 787 and Airbus A350 are expected to save much more operating cost than previous innovations.[48] This suggests that estimated industrial innovation rate p^G and, more importantly, estimated generation quality gap φ based on past data might be too conservative. In fact, estimated φ is only 12% as shown in Table 3, which implies that

[48] With the first 13 B787 delivered, its launch customer All Nippon Airways said the airplane is 21% better on fuel consumption than old models. Boeing had also claimed its 787-8 to have about 15% lower operating cost than A330-200, while Airbus predicted A350-1000 will have 25% lower fuel burn than B777-300ER.

upgrading to the next generation only increases a product's market share by 12% relative to the outside good. In contrast, industry experts predict that the new Boeing 787 and Airbus A350 will eventually drive out old generation models, indicating a much larger percentage change in market share ratio. Thus, I vary p^G and φ in all dynamic game scenarios to evaluate their impact on estimated merger efficiency. I find that:

Property 4 (The Innovation Property). *Higher innovation rate or larger generation gap increases merger efficiency for all τ. Net merger efficiency is increasing in both the innovation rate p^G and generation gap φ.*

I call the dynamic model using parameter values estimated from Section 2.4 the "base model." Comparative static analysis is then performed by varying one or more parameters of the base model and resolving the dynamic game. Figure 25 compares merger efficiency (ΔCS) across different values for the transfer rate τ in the base model, with merger efficiency in a model with doubled p^G and merger efficiency in a model with doubled φ. When p^G is doubled, consumers benefit more from the merger but not by much, and a smaller experience transfer rate shall be enough for the learning-by-doing effect to offset the market power effect ($\Delta CS = 0$). It is probably because higher p^G implies more frequent generation upgrades and setbacks in experience, favoring a more concentrated market that accumulates experience faster. Furthermore, a doubled φ generates larger consumer benefit than the doubled p^G does. When MD-11 was active, it was not upgraded to a new generation because it was expected to stop production in the long-run. Therefore, production of MD-11 leads to lower consumer surplus under the no-merger scenario. A larger generation quality gap φ magnifies the loss from MD-11 production, indicating higher consumer welfare for the merger scenario. Figure 26 plots the net merger consumer surplus ΔCS at $\tau = 1$ as a function of φ for p^G and doubled p^G. ΔCS is found to be increasing in both in p^G and φ. Thus, if the magnitude or rate of innovation is larger in the future, the merger would be more consumer beneficial.

The analysis here also provides a sensitivity check of the consumer welfare effect of the merger with respect to innovation rate p^G and generation gap φ. In Figure 25, the difference between ΔCS curve of the base model and that of the "doubled p^G" model is relatively small while net consumer surplus at $\tau = 1$ for the "doubled φ" model is more than three times larger than that for the base model. Thus, merger efficiency calculated in this paper would be too conservative if one were to believe that generation gaps should be much larger in the future.

2.6 Summary

In summarizing this paper's contribution, I will first describe the innovation in terms of model and methods and then describe the policy contribution with respect to evaluating the 1997 Boeing-McDonnell Douglas merger.

A dynamic oligopoly model is constructed that allows for multi-product firms, learning-by-doing and endogenous quality improvements. Allowing for both the evolution of cost and product quality is, to my knowledge, new to the literature on dynamic oligopoly models. Having two product-specific dynamic states (experience and generation) that evolve at multiple and different stages creates complexity in solving the dynamic model. I find that it is helpful to distinguish state profiles at different stages. Joint probabilistic upgrading decisions for a multi-product firm could be very complicated, and I deal with this by introducing randomness in a separable term (upgrading cost) that guarantees a unique analytical solution with given expected future values.

To reduce computational burdens, I also introduced a preference rank state variable to replace the unobserved characteristics state variable for each product. The preference rank state variable is applicable to any dynamic oligopoly models, including those without learning-by-doing or innovation features. Since its size does not depend on number of products, the preference rank state variable is most powerful in reducing size of state space for dynamic games with many products where variations in unobserved characteristics are primarily induced by preference shocks over certain attributes of the products, for example, ownership.

As described, the model is applicable to many industries for which learning-by-doing and quality innovation are relevant. However, the primary purpose of the model was to evaluate the 1997 Boeing-McDonnell Douglas merger in the medium-size wide-bodied aircraft industry. I find that with complete experience transfer, the merger increases consumer surplus by $1.57 billion. Consumers are better off as long as there is at least a 20% experience transfer rate after the merger. Learning-by-doing is the major source of merger efficiency and is large enough to cover the detrimental market power effect of about $20 billion. The merger's impacts on both consumer welfare and market structure are intermediate; it only accelerates experience accumulation towards the steady state and there is no-merger effect in the long-run. Comparative statics suggest that if future generation gaps were to be larger, merger efficiency would be even greater. Differences in net consumer surplus between the dynamic model and a static model suggest potential caveats in traditional static analysis in antitrust practices.

While the primary purpose of the model was to empirically investigate the aircraft manufacturing industry, the model is applicable to many industries for which learning-by-doing and quality innovation are relevant. For example, the model can be modified to examine the potential impact of the recently turned-down takeover of Seagate by Western Digital in the hard disc drive industry. More generally, the model and methods developed here may prove useful for gaining an improved empirical assessment of the significance of dynamic efficiencies from mergers.

References

ANDREWS, D. W. K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59(3), 817–858.

BENKARD, C. (2000): "Learning and Forgetting: The Dynamics of Aircraft Production," *American Economic Review*, 90(4), 1034–1054.

——— (2004): "A Dynamic Analysis of the Market for Wide-Bodied Commercial Aircraft," *Review of Economic Studies*, 71, 581–611.

BENKARD, C., A. BODOH-CREED, AND J. LAZAREV (2010): "Simulating the dynamic effects of horizontal mergers: US airlines," Discussion paper, Citeseer.

BESANKO, D., U. DORASZELSKI, Y. KRYUKOV, AND M. SATTERTHWAITE (2010): "Learning-by-doing, Organizational Forgetting, and Industry Dynamics," *Econometrica*, 78(2), 453–508.

CHANDRA, A., AND D. STAIGER (2007): "Productivity spillovers in healthcare: evidence from the treatment of heart attacks," *Journal of Political Economy*, 115, 103.

CHEN, J. (2009): "The Effects of Mergers with Dynamic Capacity Accumulation," *International Journal of Industrial Organization*, 27(1), 92–109.

DORASZELSKI, U., AND A. PAKES (2007): "A Framework for Applied Dynamic Analysis in IO," *Handbook of Industrial Organization*, 3, 1887–1966.

ERICSON, R., AND A. PAKES (1995): "Markov-Perfect Industry Dynamics: A Framework for Empirical Work," *Review of Economic Studies*, 62(1), 53–82.

GOETTLER, R. L., AND B. R. GORDON (2011): "Does AMD Spur Intel to Innovate More?," *Journal of Political Economy*, 119(6), 1141–1200.

GOWRISANKARAN, G. (1999): "A dynamic model of endogenous horizontal mergers," *RAND Journal of Economics*, 30(1), 56–83.

Gowrisankaran, G., V. Ho, and R. J. Town (2006): "Causality, Learning and Forgetting in Surgery," Working Paper.

Grossman, M. G., and E. Helpman (1995): "Chapter 25 Technology and trade," *Handbook of International Economics*, pp. 1279–1337.

Hashmi, A. R., and J. V. Biesebroeck (2010): "Market Structure and Innovation: A Dynamic Analysis of the Global Automobile Industry," Working Paper.

Irwin, A. D., and P. J. Klenow (1994): "Learning-by-Doing Spillovers in the Semiconductor Industry," *Journal of Political Economy*, 102(6), 1200–1227.

Jeziorski, P. (2011a): "Estimation of cost synergies from mergers without cost data: Application to U.S. radio," Working Paper.

——— (2011b): "Merger enforcement in two-sided markets," Working Paper.

Levitt, S. D., J. A. List, and C. Syverson (2012): "Toward an Understanding of Learning by Doing: Evidence From an Automobile Assembly Plant," NBER Working Paper.

Liu, H. (2010): "Dynamics of Pricing in the Video Game Console Market: Skimming or Penetration?," *Journal of Marketing Research*, 47(3), 428–443.

Niosi, J., and M. Zhegu (2010): "Multinational Corporations, Value Chains and Knowledge Spillovers in the Global Aircraft Industry," *International Journal of Institutions and Economies*, 2(2), 109–141.

Nocke, V., and M. D. Whinston (2010): "Dynamic Merger Review," *Journal of Political Economy*, 118, 1200–1251.

Nosko, C. (2010): "Competition and Quality Choice in the CPU Market," Working Paper.

Ohashi, H. (2005): "Learning by doing, export subsidies, and industry growth: Japanese steel in the 1950s and 1960s," *Journal of International Economics*, 66(2), 297–323.

PAKES, A., AND P. MCGUIRE (1994): "Computing Markov-Perfect Nash Equilibria: Numerical Implications of a Dynamic Differentiated Product Model," *RAND Journal of Economics*, 25(4), 555–589.

SCHWOON, M. (2008): "Learning by doing, learning spillovers and the diffusion of fuel cell vehicles," *Simulation Modelling Practice and Theory*, 16(9), 1463–1476.

SMALL, K. A., AND H. S. ROSEN (1981): "Applied Welfare Economics with Discrete Choice Models," *Econometrica*, 49(1), 105–130.

STAHL, J. (2009): "A dynamic analysis of consolidation in the broadcast television industry," *Finance and Economics Discussion Series*.

THOMPSON, P. (2001): "How Much Did the Liberty Shipbuilders Learn? New Evidence for an Old Case Study," *Journal of Political Economy*, 109(1), 103–137.

——— (2007): "How Much Did the Liberty Shipbuilders Forget?," *Management Science*, 53(6), 908–918.

THORNTON, R., AND P. THOMPSON (2001): "Learning from experience and learning from others: An exploration of learning and spillovers in wartime shipbuilding," *The American Economic Review*, 91(5), 1350–1368.

TRAJTENBERG, M. (1989): "The Welfare Analysis of Product Innovations, with an Application to Computed Tomography Scanners," *Journal of Political Economy*, 97(2), 444–479.

WRIGHT, T. (1936): "Factors affecting the cost of airplanes," *Journal of Aeronautical Sciences*.

2.7 Appendix

2.7.1 Treating MD-11 as a Within Firm Product for L-1011

There are 5 aircraft with 11 submodels that have overlaps in production life with Lockheed L-1011. To estimate Equation (12), I need to observe different families of aircraft within the Lockheed corporation. However, L-1011 is the only wide-bodied commercial aircraft that Lockheed had ever produced. Hence, I make a further assumption that DC-10 of McDonnell Douglas can be treated as a within-firm product for Lockheed L-1011.[49] Seemingly a strong assumption, using an outside product as an inside product tends to under-estimate the within firm spillover effect, which would lead to conservative estimate of merger efficiency. However, the under-estimation problem could be (partially) relieved considering the following arguments. First, L-1011 and DC-10 are probably the pair of most similar aircraft in the entire history of wide-bodied aircraft industry. Both aircraft are fitted with three high-bypass turbofan engines, seat around 300 passengers, and have about the same fuselage diameter and exactly the same wingspan. They are also very similar in many detailed aspects that could not all be covered by the product difference function in Equation (12). Hence, with respect to similarity, DC-10 should have the highest spillover on L-1011, offsetting part of the under-estimation. Second, DC-10 was put into production about a year before L-1011 and stayed in lead during their whole production histories. Combining with the similarities, this implies that Lockheed would enjoy a followers advantage in production techniques and benefit more than average cases in experience sharing from the production of DC-10. Third and probably the most important reasoning, the plants for producing L-1011 and DC-10 both sat in the Los Angeles area while the plants of Boeing and Airbus were far away in Seattle and Europe, respectively. Lockheed's plant was in Burbank, California while Douglas manufactured in Long Beach, California. The two cities are on the opposite side of Los Angeles City with about 30 miles between them. Producing together in the Los Angeles County since the 1920's implies an unique close connection between workers of

[49] An alternative choice is to use Lockheed's military aircraft C-5A Galaxy produced during the period of 1968-1973. C-5A Galaxy has the same number of engines and very similar plane size, range, and other characteristics as Boeing 747. Problems in using C-5A lie in the differences between military and commercial aircraft. For example, C-5A Galaxy does not have a corresponding number of seats characteristic as it is an airlift. In addition, C-5A had shorter overlapping production periods with L-1011 than DC-10 did.

the two companies. When there are layoffs and other mobilities, work force can easily shift from one firm to the other. Experience can be shared in the unions or even in the bars. Since within-firm spillover mainly comes from workforce shifts and experience sharing, the spillover effect between the two plants should be closer to the level of within-firm spillovers.

2.7.2 Alternative Modeling of Experience Transition

The expected value function and therefore the right hand side of the Bellman equation (Equation (5)) is not continuously differentiable in quantities according to the numerical transition rule in Equation (16). This non-differentiability makes it hard to solve for optimal quantities in numerical computation. Here I propose an alternative modeling of experience transition to smooth the expected value function.[50]

The problem of non-differentiability comes from the part that Equation (16) restricts transitions to closest lower and upper bounds/grids only while next period experience could span the entire space for theoretical experience transition (Equation (12)). To overcome this problem, a first thought is to add in a random shock to the process as

$$E_{t+1} = E^k = argmin| \left(E^*_{t+1} + \varepsilon \right) - E^k |. \tag{19}$$

Thus, the discrete experience state E_{t+1} is chosen by using E^*_{t+1} plus a random normal draw ϵ and then rounding to the nearest grid. For example, if $\epsilon \sim N(0, \sigma)$, then no matter what E_t and q_t are, $\left(E^*_{t+1} + \epsilon \right)$ could be any real number and E_{t+1} has positive probability going to any grid. This solves the non-differentiability problem but creates incorrect expectation as expectation of E_{t+1} would then not be equal to E^*_{t+1}. To fix this problem, I turn to a mixture of Equation (16) and Equation (19). The mixture is accomplished in two steps. First, I define $\widetilde{E}$ as

$$\widetilde{E} = E^*(q) + \nu, \tag{20}$$

where $\nu \sim f(\nu)$, with $E(\nu) = 0$ and CDF $F(\nu)$. Then I replace E^*_{t+1} with $\widetilde{E}$ in Equation (16). It is easy to see that $\sum_k \Pr\left(E^k\right) V^k$ is continuously differentiable under this alternative transition rule. I then show that expectation of E_{t+1} equals E^*_{t+1}. Let N be the number of

[50] I am greatly indebt to C. Lanier Benkard for suggesting this alternative modeling idea.

grids for E. Then,

$$\begin{aligned}
\Pr\left(E^1\right) &= \Pr\left(E = E^1 | \widetilde{E} < E^1\right) \Pr\left(\widetilde{E} < E^1\right) \\
&\quad + \Pr\left(E = E^1 | E^1 < \widetilde{E} < E^2\right) \Pr\left(E^1 < \widetilde{E} < E^2\right) \\
&= 1 \cdot F\left(E^1 - E^*\right) + \int_{E^1 - E^*}^{E^2 - E^*} \left(1 - \frac{E^* + \nu - E^1}{E^2 - E^1}\right) f(\nu)\, d\nu
\end{aligned} \tag{21}$$

Similarly, for $k = 2, 3, ..., (N-1)$

$$\begin{aligned}
\Pr\left(E^k\right) &= \Pr\left(E = E^k | E^{k-1} < \widetilde{E} < E^k\right) \Pr\left(E^{k-1} < \widetilde{E} < E^k\right) \\
&\quad + \Pr\left(E = E^k | E^k < \widetilde{E} < E^{k+1}\right) \Pr\left(E^k < \widetilde{E} < E^{k+1}\right) \\
&= \int_{E^{k-1} - E^*}^{E^k - E^*} \left(\frac{E^* + \nu - E^{k-1}}{E^k - E^{k-1}}\right) f(\nu)\, d\nu + \int_{E^k - E^*}^{E^{k+1} - E^*} \left(1 - \frac{E^* + \nu - E^k}{E^{k+1} - E^k}\right) f(\nu)\, d\nu
\end{aligned} \tag{22}$$

and

$$\begin{aligned}
\Pr\left(E^N\right) &= \Pr\left(E = E^N | \widetilde{E} > E^N\right) \Pr\left(\widetilde{E} > E^N\right) \\
&\quad + \Pr\left(E = E^N | E^{N-1} < \widetilde{E} < E^N\right) \Pr\left(E^{N-1} < \widetilde{E} < E^N\right) \\
&= 1 \cdot \left(1 - F\left(E^N - E^*\right)\right) + \int_{E^{N-1} - E^*}^{E^N - E^*} \left(\frac{E^* + \nu - E^{N-1}}{E^N - E^{N-1}}\right) f(\nu)\, d\nu
\end{aligned} \tag{23}$$

Then

$$\begin{aligned}
& \sum_k \Pr\left(E^k\right) E^k && (24)\\
= \; & E^1 \cdot F\left(E^1 - E^*\right) + E^N \cdot \left(1 - F\left(E^N - E^*\right)\right) \\
& + \sum_{k=1}^{N-1} E^k \int_{E^k - E^*}^{E^{k+1} - E^*} 1 \cdot f(\nu)\, d\nu + \int_{E^1 - E^*}^{E^N - E^*} \left(E^* + \nu\right) \cdot f(\nu)\, d\nu \\
& - \sum_{k=1}^{N-1} E^k \int_{E^k - E^*}^{E^{k+1} - E^*} 1 \cdot f(\nu)\, d\nu \\
= \; & E^1 \cdot F\left(E^1 - E^*\right) + E^N \cdot \left(1 - F\left(E^N - E^*\right)\right) \\
& + E^* \cdot \left[F\left(E^N - E^*\right) - F\left(E^1 - E^*\right)\right] + \int_{E^1 - E^*}^{E^N - E^*} \nu \cdot f(\nu)\, d\nu \\
= \; & \left(E^1 - E^*\right) \cdot F\left(E^1 - E^*\right) + \left(E^N - E^*\right) \cdot \left(1 - F\left(E^N - E^*\right)\right) \\
& + E^* + \int_{E^1 - E^*}^{E^N - E^*} \nu \cdot f(\nu)\, d\nu \\
= \; & E^* + \int_{-\infty}^{E^1 - E^*} \left(E^1 - E^* - \nu\right) f(\nu)\, d\nu + \int_{E^N - E^*}^{\infty} \left(E^N - E^* - \nu\right) f(\nu)\, d\nu \\
\neq \; & E^*
\end{aligned}$$

where the last equation used the fact $E(\nu) = 0$.

However, note that if I restrict the domain of ν to be $\left(E^1 - E^*, E^N - E^*\right)$, then

$$\begin{aligned}
\int_{E^1 - E^*}^{E^N - E^*} \nu \cdot f(\nu)\, d\nu &= E(\nu) = 0; && (25)\\
F\left(E^1 - E^*\right) &= 0; \\
F\left(E^N - E^*\right) &= 1; \\
E^* \cdot \int_{E^1 - E^*}^{E^N - E^*} f(\nu)\, d\nu &= E^* \cdot 1 = E^*. && (26)
\end{aligned}$$

Hence,

$$\sum_k \Pr\left(E^k\right) E^k = E^*.$$

Note that with Equation (12), next period experience is bounded below by δE_{t-1} but not so under this alternative transition. This problem could be fixed by defining lower bound of next period E to be δE_{t-1}. However, this will further complicates the function form to achieve correct expectation.

There are several problems with this alternative modeling though. First, it introduces non-negligible computational burden by employing a more complicated function form. Second, it might create too much variation of experience evolution than the data does. under this alternative transition rule, it is possible for experience to jump from the lowest state to the highest state even when $q = 0$ (or from the highest state to the lowest state even when q is very large). This would cause problems for some simulation paths. However, it won't be a problem for expected discounted CS since probability of such events will be very small.

2.7.3 Proof on Uniqueness of $Pr_i^{U_i}$ for Given $EV_i^{U_i}$

In Equation (3), $EV_i^{U_i}$ is not a function of any c_j for given U_i. Hence, the difference between $EV_i^{U_i}$ and $EV_i^{U_i'}$ is monotonic in one or more c_j. For given C_i^G, $EV_i^{U_i} - C_i^G$ is simply a vector of $2^{\mathbb{J}_i}$ numbers and we can always find the largest number. Thus, C_i^G divides the $\mathbb{J}_i$-*dimension* Euclidean space $[C^d, C^u]^{\mathbb{J}_i}$ into $2^{\mathbb{J}_i}$ areas in a unique way[51]. $Pr_i^{U_i}$ then is unique and equals the proportion of areas of the Euclidean space $[C^d, C^u]^{\mathbb{J}_i}$.[52]

I give an example on finding the unique solution of upgrade probabilities for a two-product firm below. Models with firms that have more than 2 products can be solved similarly. In the example, for subscript ij, i and j denotes product 1 and 2 of the firm, respectively. i (or j) equals 1 (or 0) if product i (or j) is upgraded (not upgraded). There are four possible choices for the firm:

$$U_{11}, U_{10}, U_{01}, U_{00}.$$

The problem can be summarized as solving

$$P_{11}, P_{10}, P_{01}, P_{00},$$

for given continuation values

$$EV_{11}, EV_{10}, EV_{01}, EV_{00}.$$

[51]Strictly speaking, there might be less than $2^{\mathbb{J}_i}$ areas since some U_i might be so undesirable that it is never chosen for any $c_j \in [C^d, C^u]^{\mathbb{J}_i}$

[52]Note that the upgrade probabilities $Pr_i^{U_i}$ are continuous but generally not differentiable (at boundaries between cases) in $EV_i^{U_i}$.

The net value from each choice is then:

$$\begin{aligned} V_{11} &= EV_{11} - c_1 - c_2; \\ V_{10} &= EV_{10} - c_1; \\ V_{01} &= EV_{01} - c_2; \\ V_{00} &= EV_{00}. \end{aligned}$$

Comparing V_{ij} defines the following six lines in the c_1-c_2 plane that marks the boundaries between V_{ij}.

$$\begin{array}{lll} L_1: & EV_{00} - EV_{10} + c_1 = 0 & \text{(vertical)} \\ L_2: & EV_{01} - EV_{11} + c_1 = 0 & \text{(vertical)} \\ L_3: & EV_{00} - EV_{01} + c_2 = 0 & \text{(horizontal)} \\ L_4: & EV_{10} - EV_{11} + c_2 = 0 & \text{(horizontal)} \\ L_5: & EV_{10} - EV_{01} + c_2 - c_1 = 0 & (45^\circ) \\ L_6: & EV_{00} - EV_{11} + c_1 + c_2 = 0 & (-45^\circ) \end{array}$$

It is important to notice that the six boundary lines satisfy the following relationship:

$$\begin{aligned} L_1 + L_4 &= L_6 \\ L_2 + L_3 &= L_6 \\ L_4 - L_2 &= L_5 \\ L_3 - L_1 &= L_5 \end{aligned}$$

In addition, upgrade choices are indicated in the c_1-c_2 plane by areas defined by the six boundary lines as follows:

$$\begin{array}{lcl} U_{11} & : & \text{left of } L_2\text{, under } L_4\text{, under } L_6; \\ U_{10} & : & \text{left of } L_1\text{, above } L_4\text{, above } L_5; \\ U_{01} & : & \text{right of } L_2\text{, under } L_3\text{, under } L_5; \\ U_{00} & : & \text{right of } L_1\text{, above } L_3\text{, above } L_6. \end{array}$$

Specific divisions of the c_1-c_2 plane can be divided into two major cases, with $EV_{00}+EV_{11}-EV_{01}-EV_{10}\geq 0$ and $EV_{00}+EV_{11}-EV_{01}-EV_{10}<0$ demonstrated, respectively, in Figure 27 and 28.

Finally, note that C^d and C^u define a square box, $\left[C^d,C^u\right]\times\left[C^d,C^u\right]$, in the c_1-c_2 plane. The upgrade probabilities P_{ij} then equal their corresponding percentage of areas within the square box $\left[C^d,C^u\right]\times\left[C^d,C^u\right]$. Specifically, depending on the size of the square box and its relative position to the six boundary lines, the division of the square box can be categorized into the following 24 cases. The upgrade probabilities P_{ij} are solved according to the following formulas in each case:

- For $EV_{00}+EV_{11}-EV_{01}-EV_{10}\geq 0$
  - Case I(i)
    - Condition:

$$\begin{aligned} C^d &> EV_{11}-EV_{10}; \\ C^u &\leq EV_{10}-EV_{00} \end{aligned}$$

    - Probability:

$$P_{10}=1$$

  - Case I(ii)
    - Condition:

$$\begin{aligned} C^d &> EV_{11}-EV_{10}; \\ C^d &\leq EV_{10}-EV_{00}<C^u \end{aligned}$$

    - Probability:

$$P_{10}=\frac{EV_{10}-EV_{00}-C^d}{C^u-C^d}; P_{00}=1-P_{10}$$

  - Case I(iii)

* Condition:

$$\begin{aligned} EV_{00} - EV_{11} + 2C^d &\geq 0; \\ EV_{10} - EV_{00} &< C^d; \\ EV_{01} - EV_{00} &< C^d \end{aligned}$$

* Probability:

$$P_{00} = 1$$

- Case I(iv)

  * Condition:

$$\begin{aligned} C^d &\leq EV_{11} - EV_{10} < C^u; \\ C^u &\leq EV_{10} - EV_{00} \end{aligned}$$

  * Probability:

$$P_{11} = \frac{EV_{11} - EV_{10} - C^d}{C^u - C^d}; P_{10} = 1 - P_{11}$$

- Case I(v)

  * Condition:

$$\begin{aligned} C^d &\leq EV_{10} - EV_{00} < C^u; \\ C^d &\leq EV_{11} - EV_{10} < C^u; \end{aligned}$$

  * Probability:

$$\begin{aligned} P_{10} &= \frac{EV_{10} - EV_{00} - C^d}{C^u - C^d} \cdot \frac{C^u - (EV_{11} - EV_{10})}{C^u - C^d} \\ S_{11} &= \frac{EV_{10} - EV_{00} - C^d}{C^u - C^d} \cdot \frac{EV_{11} - EV_{10} - C^d}{C^u - C^d} \\ S_{00} &= \frac{C^u - (EV_{10} - EV_{00})}{C^u - C^d} \cdot \frac{C^u - (EV_{11} - EV_{10})}{C^u - C^d} \\ S_{joint} &= 1 - P_{10} - S_{11} - S_{00} \end{aligned}$$

if,

$$EV_{00} - EV_{11} + C^u + C^d \geq 0$$

$$\begin{aligned} S_{joint\_11} &= \frac{\left(EV_{11} - EV_{10} - C^d\right)^2}{2\left(C^u - C^d\right)^2} \\ P_{11} &= S_{11} + S_{joint\_11} \\ P_{00} &= S_{00} + S_{joint} - S_{joint\_11} \end{aligned}$$

eles if,

$$EV_{00} - EV_{11} + C^u + C^d < 0$$

$$\begin{aligned} S_{joint\_00} &= \frac{\left(EV_{10} - EV_{00} - C^u\right)^2}{2\left(C^u - C^d\right)^2} \\ P_{11} &= S_{11} + S_{joint} - S_{joint\_00} \\ P_{00} &= S_{00} + S_{joint\_00} \end{aligned}$$

- Case I(vi)
  * Condition:

$$\begin{aligned} EV_{00} - EV_{11} + 2C^u &< 0; \\ EV_{11} - EV_{10} &\geq C^u; \\ EV_{11} - EV_{01} &\geq C^u \end{aligned}$$

  * Probability:

$$P_{11} = 1$$

- Case I(vii)

* Condition:

$$\begin{aligned} C^d &> EV_{10} - EV_{00}; \\ C^d &> EV_{01} - EV_{00}; \\ C^u &\leq EV_{11} - EV_{01}; \\ C^u &\leq EV_{11} - EV_{10} \end{aligned}$$

* Probability:

$$\begin{aligned} &P_{11} = \tfrac{\left(EV_{11}-EV_{00}-2C^d\right)^2}{2\left(C^u-C^d\right)^2}; P_{00} = 1 - P_{11} \quad \text{if } EV_{00} - EV_{11} + C^d + C^u \geq 0 \\ &P_{00} = \tfrac{(2C^u-EV_{11}+EV_{00})^2}{2\left(C^u-C^d\right)^2}; P_{11} = 1 - P_{00} \quad \text{otherwise} \end{aligned}$$

- Case I(viii)

* Condition:

$$\begin{aligned} C^d &\leq EV_{11} - EV_{01} < C^u; \\ C^d &\leq EV_{01} - EV_{00} < C^u \end{aligned}$$

* Probability:

$$\begin{aligned} P_{01} &= \frac{EV_{01} - EV_{00} - C^d}{C^u - C^d} \cdot \frac{C^u - (EV_{11} - EV_{01})}{C^u - C^d} \\ S_{11} &= \frac{EV_{01} - EV_{00} - C^d}{C^u - C^d} \cdot \frac{EV_{11} - EV_{01} - C^d}{C^u - C^d} \\ S_{00} &= \frac{C^u - (EV_{01} - EV_{00})}{C^u - C^d} \cdot \frac{C^u - (EV_{11} - EV_{01})}{C^u - C^d} \\ S_{joint} &= 1 - P_{01} - S_{11} - S_{00} \end{aligned}$$

if,

$$EV_{00} - EV_{11} + C^d + C^u \geq 0$$

$$
\begin{aligned}
S_{joint\_11} &= \frac{\left(EV_{11} - EV_{01} - C^d\right)^2}{2\left(C^u - C^d\right)^2} \\
P_{11} &= S_{11} + S_{joint\_11} \\
P_{00} &= S_{00} + S_{joint} - S_{joint\_11}
\end{aligned}
$$

eles if,

$$EV_{00} - EV_{11} + C^d + C^u < 0$$

$$
\begin{aligned}
S_{joint\_00} &= \frac{\left(EV_{01} - EV_{00} - C^u\right)^2}{2\left(C^u - C^d\right)^2} \\
P_{11} &= S_{11} + S_{joint} - S_{joint\_00} \\
P_{00} &= S_{00} + S_{joint\_00}
\end{aligned}
$$

– Case I(ix)

  * Condition:

$$
\begin{aligned}
C^d &> EV_{11} - EV_{01}; \\
C^d &\leq EV_{01} - EV_{00} < C^u
\end{aligned}
$$

  * Probability:

$$P_{01} = \frac{EV_{01} - EV_{00} - C^d}{C^u - C^d}; P_{00} = 1 - P_{01}$$

– Case I(x)

  * Condition:

$$
\begin{aligned}
C^d &\leq EV_{11} - EV_{01} < C^u; \\
C^u &\leq EV_{01} - EV_{00}
\end{aligned}
$$

  * Probability:

$$P_{11} = \frac{EV_{11} - EV_{01} - C^d}{C^u - C^d}; P_{01} = 1 - P_{11}$$

– Case I(xi)

* Condition:

$$\begin{aligned} C^d &> EV_{11} - EV_{01}; \\ C^u &\leq EV_{01} - EV_{00} \end{aligned}$$

* Probability:

$$P_{01} = 1$$

- Case I(xii)

  * Condition:

$$\begin{aligned} EV_{11} - EV_{10} &< C^u \\ EV_{01} - EV_{00} &\geq C^d \\ EV_{11} - EV_{01} &< C^u \\ EV_{10} - EV_{00} &\geq C^d \end{aligned}$$

  * Probability:

$$\begin{aligned} P_{01} &= \frac{EV_{10} - EV_{00} - C^d}{C^u - C^d} \cdot \frac{C^u - (EV_{11} - EV_{10})}{C^u - C^d} \\ P_{10} &= \frac{EV_{01} - EV_{00} - C^d}{C^u - C^d} \cdot \frac{C^u - (EV_{11} - EV_{01})}{C^u - C^d} \\ S_{11} &= \frac{EV_{11} - EV_{01} - C^d}{C^u - C^d} \cdot \frac{EV_{11} - EV_{10} - C^d}{C^u - C^d} \\ S_{00} &= \frac{C^u - (EV_{01} - EV_{00})}{C^u - C^d} \cdot \frac{C^u - (EV_{10} - EV_{00})}{C^u - C^d} \\ S_{joint} &= P_{10} + P_{01} + S_{11} + S_{00} - 1 \end{aligned}$$

  then

$$\begin{aligned} P_{11} &= S_{11} - \frac{1}{2} S_{joint} \\ P_{00} &= S_{00} - \frac{1}{2} S_{joint} \end{aligned}$$

- For $EV_{00} + EV_{11} - EV_{01} - EV_{10} < 0$,

- Case II(i)
  - Condition:

$$\begin{aligned} C^d &> EV_{10} - EV_{00}; \\ C^d &> EV_{01} - EV_{00} \end{aligned}$$

  - Probability:

$$P_{00} = 1$$

- Case II(ii)
  - Condition:

$$\begin{aligned} C^d &> EV_{01} - EV_{00}; \\ C^d &\leq EV_{10} - EV_{00} < C^u \end{aligned}$$

  - Probability:

$$P_{10} = \frac{EV_{10} - EV_{00} - C^d}{C^u - C^d}; P_{00} = 1 - P_{10}$$

- Case II(iii)
  - Condition:

$$\begin{aligned} EV_{10} - EV_{01} + C^d - C^u &\geq 0; \\ EV_{10} - EV_{00} &\geq C^u; \\ EV_{11} - EV_{10} &< C^d \end{aligned}$$

  - Probability:

$$P_{10} = 1$$

- Case II(iv)

* Condition:

$$\begin{aligned} C^d &\leq EV_{01} - EV_{00} < C^u; \\ C^d &> EV_{10} - EV_{00} \end{aligned}$$

* Probability:

$$P_{01} = \frac{EV_{01} - EV_{00} - C^d}{C^u - C^d}; P_{00} = 1 - P_{01}$$

- Case II(v)

* Condition:

$$\begin{aligned} C^d &\leq EV_{10} - EV_{00} < C^u; \\ C^d &\leq EV_{01} - EV_{00} < C^u \end{aligned}$$

* Probability:

$$\begin{aligned} P_{00} &= \frac{C^u - (EV_{10} - EV_{00})}{C^u - C^d} \cdot \frac{C^u - (EV_{01} - EV_{00})}{C^u - C^d} \\ S_{10} &= \frac{EV_{10} - EV_{00} - C^d}{C^u - C^d} \cdot \frac{C^u - (EV_{01} - EV_{00})}{C^u - C^d} \\ S_{01} &= \frac{C^u - (EV_{10} - EV_{00})}{C^u - C^d} \cdot \frac{EV_{01} - EV_{00} - C^d}{C^u - C^d} \\ S_{joint} &= 1 - P_{00} - S_{10} - S_{01} \end{aligned}$$

if,

$$EV_{10} - EV_{01} + C^d - C^d = EV_{10} - EV_{01} \geq 0$$

$$\begin{aligned} S_{joint\_01} &= \frac{\left(EV_{01} - EV_{00} - C^d\right)^2}{2\left(C^u - C^d\right)^2} \\ P_{01} &= S_{01} + S_{joint\_01} \\ P_{10} &= S_{10} + S_{joint} - S_{joint\_01} \end{aligned}$$

eles if,

$$EV_{10} - EV_{01} + C^d - C^d = EV_{10} - EV_{01} < 0$$

$$
\begin{aligned}
S_{joint\_10} &= \frac{\left(EV_{10} - EV_{00} - C^d\right)^2}{2\left(C^u - C^d\right)^2} \\
P_{01} &= S_{01} + S_{joint} - S_{joint\_10} \\
P_{10} &= S_{10} + S_{joint\_10}
\end{aligned}
$$

- Case II(vi)
  - Condition:

$$
\begin{aligned}
EV_{10} - EV_{01} + C^u - C^d &< 0; \\
EV_{01} - EV_{00} &\geq C^u; \\
EV_{11} - EV_{01} &< C^d
\end{aligned}
$$

  - Probability:

$$P_{01} = 1$$

- Case II(vii)
  - Condition:

$$
\begin{aligned}
C^d &> EV_{11} - EV_{01}; \\
C^d &> EV_{11} - EV_{10}; \\
C^u &\leq EV_{10} - EV_{00}; \\
C^u &\leq EV_{01} - EV_{00}
\end{aligned}
$$

  - Probability:

$$
\begin{array}{ll}
P_{01} = \frac{\left(C^u - (EV_{10} - EV_{01}) - C^d\right)^2}{2\left(C^u - C^d\right)^2}; P_{10} = 1 - P_{01} & \text{if } EV_{10} - EV_{01} \geq 0 \\
P_{10} = \frac{\left(C^u + (EV_{10} - EV_{01}) - C^d\right)^2}{2\left(C^u - C^d\right)^2}; P_{01} = 1 - P_{10} & \text{otherwise}
\end{array}
$$

- Case II(viii)

* Condition:

$$\begin{aligned} C^d &\leq EV_{11} - EV_{01} < C^u; \\ C^d &\leq EV_{11} - EV_{10} < C^u \end{aligned}$$

* Probability:

$$\begin{aligned} P_{11} &= \frac{EV_{11} - EV_{10} - C^d}{C^u - C^d} \cdot \frac{EV_{11} - EV_{01} - C^d}{C^u - C^d} \\ S_{10} &= \frac{EV_{11} - EV_{01} - C^d}{C^u - C^d} \cdot \frac{C^u - (EV_{11} - EV_{10})}{C^u - C^d} \\ S_{01} &= \frac{C^u - (EV_{11} - EV_{01})}{C^u - C^d} \cdot \frac{EV_{11} - EV_{10} - C^d}{C^u - C^d} \\ S_{joint} &= 1 - P_{11} - S_{10} - S_{01} \end{aligned}$$

if,

$$EV_{10} - EV_{01} + C^u - C^u = EV_{10} - EV_{01} \geq 0$$

$$\begin{aligned} S_{joint\_01} &= \frac{(EV_{11} - EV_{01} - C^u)^2}{2\left(C^u - C^d\right)^2} \\ P_{01} &= S_{01} + S_{joint\_01} \\ P_{10} &= S_{10} + S_{joint} - S_{joint\_01} \end{aligned}$$

eles if,

$$EV_{10} - EV_{01} + C^u - C^u = EV_{10} - EV_{01} < 0$$

$$\begin{aligned} S_{joint\_10} &= \frac{(EV_{11} - EV_{10} - C^u)^2}{2\left(C^u - C^d\right)^2} \\ P_{01} &= S_{01} + S_{joint} - S_{joint\_10} \\ P_{10} &= S_{10} + S_{joint\_10} \end{aligned}$$

– Case II(ix)

* Condition:

$$\begin{aligned} C^u &\leq EV_{11} - EV_{01}; \\ C^d &\leq EV_{11} - EV_{10} < C^u \end{aligned}$$

* Probability:

$$P_{11} = \frac{EV_{11} - EV_{10} - C^d}{C^u - C^d}; P_{10} = 1 - P_{11}$$

- Case II(x)
  * Condition:

$$\begin{aligned} C^u &\leq EV_{11} - EV_{10}; \\ C^d &\leq EV_{11} - EV_{01} < C^u \end{aligned}$$

  * Probability:

$$P_{11} = \frac{EV_{11} - EV_{01} - C^d}{C^u - C^d}; P_{01} = 1 - P_{11}$$

- Case II(xi)
  * Condition:

$$\begin{aligned} C^u &\leq EV_{11} - EV_{10}; \\ C^u &\leq EV_{11} - EV_{01} \end{aligned}$$

  * Probability:

$$P_{11} = 1$$

- Case II(xii)

- Condition:

$$
\begin{aligned}
EV_{10} - EV_{00} &< C^u \\
EV_{11} - EV_{10} &\geq C^d \\
EV_{01} - EV_{00} &< C^u \\
EV_{11} - EV_{01} &\geq C^d
\end{aligned}
$$

- Probability:

$$
\begin{aligned}
P_{11} &= \frac{EV_{11} - EV_{01} - C^d}{C^u - C^d} \cdot \frac{EV_{11} - EV_{10} - C^d}{C^u - C^d} \\
P_{00} &= \frac{C^u - (EV_{10} - EV_{00})}{C^u - C^d} \cdot \frac{C^u - (EV_{01} - EV_{00})}{C^u - C^d} \\
S_{10} &= \frac{EV_{10} - EV_{00} - C^d}{C^u - C^d} \cdot \frac{C^u - (EV_{11} - EV_{10})}{C^u - C^d} \\
S_{01} &= \frac{C^u - (EV_{11} - EV_{01})}{C^u - C^d} \cdot \frac{EV_{01} - EV_{00} - C^d}{C^u - C^d} \\
S_{joint} &= P_{11} + P_{00} + S_{10} + S_{01} - 1
\end{aligned}
$$

then

$$
\begin{aligned}
P_{10} &= S_{10} - \frac{1}{2} S_{joint} \\
P_{01} &= S_{01} - \frac{1}{2} S_{joint}
\end{aligned}
$$

2.7.4 Expected Value Function and State Transition

First I describe the state transition with experience state variables only. In this case, define

$$
\eta_j (h; q_j) = \left(\frac{E^*_{j,t+1}(q_j) - E_{j,d}(q_j)}{E_{j,u}(q_j) - E_{j,d}(q_j)} \right)^h \left(1 - \frac{E^*_{j,t+1}(q_j) - E_{j,d}(q_j)}{E_{j,u}(q_j) - E_{j,d}(q_j)} \right)^{1-h}
$$

where h is either 0 or 1. Then,

$$
\begin{aligned}
EV_j\left(\overrightarrow{q}\right) &= EV_j\left(E'\right) = EV_j\left(E_j^{'}; E_{-j}^{'}\right) \\
&= \eta_j\left(1\right) V_j\left(E_u; E\left[E_{-j}^{'}\left(\overrightarrow{q}\right)\right]\right) + \eta_j\left(0\right) V_j\left(E_d; E\left[E_{-j}^{'}\left(\overrightarrow{q}\right)\right]\right) \\
&= \sum_{h_1=0,1} \dots \sum_{h_k=0,1} \dots \sum_{h_J=0,1} \left(\prod_k \eta_k\left(h_k\right)\right) V_j\left(E_{1,h_1}, \dots E_{k,h_k}, \dots E_{J,h_J}\right)
\end{aligned}
\tag{27}
$$

where experience state transition $\Pr\left(E'|E,Q\right)$ is simply

$$
\Pr\left(E'|E,Q\right) = \sum_{h_1=0,1} \dots \sum_{h_k=0,1} \dots \sum_{h_J=0,1} \left(\prod_k \eta_k\left(h_k\right)\right)
$$

Then I add in market size state variable M_t. Note that given Q_t and E_t, E_{t+1} does not depend on M_t or M_{t+1}, i.e.

$$
\Pr\left(E', M'|E, M, Q\right) = \Pr\left(E'|E,Q\right) \cdot \Pr\left(M'|M\right)
$$

Then

$$
\begin{aligned}
EV_j &= \sum_{M'} \sum_{E'} V_j\left(E', M'\right) \cdot \Pr\left(E', M'|E, M, Q\right) \\
&= \sum_{M'} \left[\sum_{E'} V_j\left(E', M'\right) \Pr\left(E'|E,Q\right)\right] \cdot \Pr\left(M'|M\right) \\
&= \sum_{M'} EV_j^{M'} \cdot \Pr\left(M'|M\right)
\end{aligned}
\tag{28}
$$

where $EV_j^{M'}$ is just EV_j in Equation 27.

Similar as the market size state variable, the feature that preference rank state variables $\left(\kappa^T, \kappa^F\right)$ evolves separately from other state variables make computation easier. Since

$$
\Pr\left(E', M', \kappa^{T\prime}, \kappa^{F\prime}|E, M, \kappa^T, \kappa^F, Q\right) = \Pr\left(E'|E,Q\right) \cdot \Pr\left(M'|M\right) \cdot \Pr\left(\kappa^{T\prime}|\kappa^T\right) \cdot \Pr\left(\kappa^{F\prime}|\kappa^F\right)
$$

then

$$
\begin{aligned}
EV_j &= \sum_{\kappa^{F\prime}} \sum_{\kappa^{T\prime}} \sum_{M'} \sum_{E'} V_j\left(E', M', \kappa^{T\prime}, \kappa^{F\prime}\right) \cdot \Pr\left(E', M', \kappa^{T\prime}, \kappa^{F\prime} | E, M, \kappa^T, \kappa^F, Q\right) \quad (29)\\
&= \sum_{\kappa^{F\prime}} \sum_{\kappa^{T\prime}} \left\{ \sum_{M'} \left[\sum_{E'} V_j\left(E', M'\right) \Pr\left(E'|E, Q\right) \right] \cdot \Pr\left(M'|M\right) \right\} \cdot \Pr\left(\kappa^{T\prime}|\kappa^T\right) \cdot \Pr\left(\kappa^{F\prime}|\kappa^F\right)\\
&= \sum_{\kappa^{F\prime}} \sum_{\kappa^{T\prime}} \left\{ \sum_{M'} EV_j^{M'} \cdot \Pr\left(M'|M\right) \right\} \cdot \Pr\left(\kappa^{T\prime}|\kappa^T\right) \cdot \Pr\left(\kappa^{F\prime}|\kappa^F\right)\\
&= \sum_{\kappa^{F\prime}} \sum_{\kappa^{T\prime}} EV_j^{\kappa^{F\prime}, \kappa^{T\prime}} \cdot \Pr\left(\kappa^{T\prime}|\kappa^T\right) \cdot \Pr\left(\kappa^{F\prime}|\kappa^F\right)
\end{aligned}
$$

where $EV_j^{\kappa^{F\prime}, \kappa^{T\prime}}$ is just EV_j in Equation 28.

Finally, I add in generation difference state variables. I need to compute

$$\Pr\left(\widetilde{\omega}'|\widetilde{\omega}\right) = \Pr\left(\widetilde{\omega}'|\omega'\right) \cdot \Pr\left(\omega'|\omega'\right).$$

Since (M, κ^T, κ^F) evolve exogenously, I only need to specify transitions of E and G. Note that G^ω is only updated in Stage (i) and (ii) governed by Equation (1) and (2). Transition of G does not depend on transition of E but not the other way around. For given $\widetilde{\omega}$, denote G_{down} and G_{stay} as the states of G at the beginning of Stage (ii) in the next period when the event *Outside Good Generation Upgrade* took place and otherwise, respectively. For product j and given $Pr_i^{U_i}$, let Pr_j^{down}, Pr_j^{stay} and Pr_j^{up} denote probability of generation difference G_j of product j after Stage (ii) in the next period decreases, remains the same and increases, respectively. Then,

$$
\begin{aligned}
Pr_j^{down} &= p^G \cdot (Pr_i^{u_j}|G_{down}) + (1-p^G) \cdot (Pr_i^{u_j}|G_{stay})\\
Pr_j^{stay} &= (1-p^G) \cdot (Pr_i^{u_j}|G_{down})\\
Pr_j^{up} &= p^G \cdot (Pr_i^{u_j}|G_{stay}).
\end{aligned}
$$

The above equations describe the transition of G part of $\Pr\left(\widetilde{\omega}'|\omega'\right)$.[53] Given $\Pr_j^{down}$,

[53] I omit special cases of hitting the smallest and largest grids of E and G in notations here for simplicity.

the transition of E_j part is simply

$$E_j' = \begin{cases} E_j^u \text{ downgrade by } n_G & \text{with prob. } Pr_j^{down} \cdot \eta_j(1) \\ E_j^d \text{ downgrade by } n_G & \text{with prob. } Pr_j^{down} \cdot \eta_j(0) \\ E_j^u & \text{with prob. } (1 - Pr_j^{down}) \cdot \eta_j(1) \\ E_j^d & \text{with prob. } (1 - Pr_j^{down}) \cdot \eta_j(0) \end{cases}$$

Table 1: Aircraft Characteristics

Characteristics	A330	A340	B777	MD-11
Aircraft ID	1	2	3	4
first delivery	1993	1993	1995	1990
seats	270	326	325	293
range(km)	12378	14312	14067	12670
No. of engines	2	4	2	3

Table 2: Demand Function Estimates

	$R^2 =$	0.9724;	Adj. $R^2 =$	0.9711	
Variable	Estimate	S.E.	t	$p > \|t\|$	Data S.E
Constant	-3.59	0.22	-16.51	0.00	N/A
seats/100	0.11	0.06	1.91	0.06	0.36
range/10000	2.04	1.07	1.91	0.06	0.20
No. of engines	-0.07	0.02	-2.73	0.01	0.91
price/100	-0.52	0.16	-3.25	0.00	0.17
InGroup Corr. (σ)	0.98	0.04	23.81	0.00	1.13

Table 3: Demand Function Estimates (with Generation)

	$R^2 =$	0.9724;	Adj. $R^2 =$	0.9711	
Variable	Estimate	S.E.	t	$p > \|t\|$	Data S.E
Constant	-3.40	0.22	-15.46	0.00	N/A
Generation	0.12	0.06	2.02	0.05	0.49
seats/100	0.07	0.06	1.25	0.21	0.36
range/10000	0.15	0.11	1.43	0.16	0.20
No. of engines	-0.03	0.03	-0.80	0.42	0.91
price/100	-0.75	0.13	-5.90	0.00	0.17
InGroup Corr. (σ)	0.97	0.02	50.26	0.00	1.13

Table 4: Learning Curve Parameters

	Explanation	Value	Std	Value	Std
lnA	Labor Cost Intercept	9.2590	(3.2885)	9.3113	(3.1696)
γ_2	Return to Scale	0.3178	(0.5904)	0.3141	(0.5552)
γ_1	Learning Parameter	-1.1462	(0.1374)	-1.1523	(0.1275)
	Implied Learning Rate	55%		55%	
δ	Depreciation of E	0.9546	(0.0014)	0.9549	(0.0012)
θ_1	In-family Spillover	0.9999	(0.0239)	0.9742	(0.0198)
θ_2	In-firm Spillover	0.2383	(0.0029)	0.2408	(0.0001)
θ_3	Across-firm Spillover	0.0138	(0.0017)	0.0182	(0.0001)
υ_1	Seats Diff.	0.9998	(0.0037)		
υ_2	Maximum Range Diff.	0.9998	(0.0032)		

Note: Implicit learning rate is calculated as $1 - 2^{\gamma_1}$*,*
which measure percent of labor saving when experience doubles.

Table 5: Market Size and Preference Rank Parameters

$\bar{\xi}_j =$	(-1.5E-9, 6.0E-10, 0.0, -1.2E-9)
$\kappa^{T1} =$	(0.0995, 0.0590)
$\kappa^{T0} =$	(-0.0885, -0.0524)
$\kappa^{F1} =$	(-0.0156, -0.0754)
$\kappa^{F0} =$	(0.0286, 0.2074)
T transition:	$\begin{pmatrix} 0.4286 & 0.5556 \\ 0.5714 & 0.4444 \end{pmatrix}$
F transition:	$\begin{pmatrix} 0.8182 & 0.4 \\ 0.1818 & 0.6 \end{pmatrix}$
M grids:	(2823, 2966, 3100)
M transition:	$\begin{pmatrix} 0.8462 & 0.2143 & 0.0000 \\ 0.1538 & 0.7143 & 0.1429 \\ 0.0000 & 0.0714 & 0.8571 \end{pmatrix}$

Table 6: Merger Effect for *Game A* (Full Model)

	Scenario (i)		*Scenario (ii)*		*Scenario (iii)*	(i)-(iii)	
Value	(a) $\tau = 0$	(b) $\tau = 1$	(a) $\tau = 0$	(b) $\tau = 1$		(i.a)-(iii)	(i.b)-(iii)
CS	154.19	157.55	154.83	157.61	155.97	-1.78	1.57
π_{all}	30.43	36.30	27.41	33.62	26.31	4.12	9.99
TS	184.62	193.85	182.25	191.23	182.29	2.34	11.56
π_{A330}	18.27	15.13	17.87	15.17	17.05	1.22	-1.92
π_{A340}	-2.23	-2.35	-2.27	-2.35	-2.30	0.07	-0.05
π_{B777}	14.39	23.52	14.19	23.35	13.83	0.55	9.69
π_{MD11}	N/A	N/A	-2.38	-2.54	-2.27	2.27	2.27

All values are total discounted expected values in billions of 1994 U.S. dollar.

Table 7: Merger Effect for *Game B* (without Generation Upgrade)

	Scenario (i)		*Scenario (ii)*		*Scenario (iii)*	(i)-(iii)	
Value	(a) $\tau = 0$	(b) $\tau = 1$	(a) $\tau = 0$	(b) $\tau = 1$		(i.a)-(iii)	(i.b)-(iii)
CS	177.89	184.06	178.36	184.06	180.68	-2.79	3.38
π_{all}	58.99	63.64	56.09	61.02	53.98	5.01	9.66
TS	236.87	247.69	234.46	245.08	234.66	2.22	13.04
π_{A330}	24.04	20.69	23.71	20.69	22.34	1.71	-1.65
π_{A340}	-1.45	-1.97	-1.52	-1.98	-1.69	0.24	-0.28
π_{B777}	36.40	44.92	36.27	44.92	35.28	1.11	9.64
π_{MD11}	N/A	N/A	-2.36	-2.61	-1.95	1.95	1.95

All values are total discounted expected values in billions of 1994 U.S. dollar.

Table 8: Merger Effect for *Game C* (the Static Game)

Value	*Scenario (i)*	*Scenario (ii)*	*Scenario (iii)*	(i)-(iii)
CS	63.05	82.99	85.58	-22.53
π_{all}	51.91	45.77	43.68	8.23
TS	114.96	128.76	129.26	-14.30
π_{A330}	26.56	17.22	16.18	10.38
π_{A340}	23.45	15.50	14.63	8.82
π_{B777}	1.90	-0.65	-0.25	2.15
π_{MD11}	N/A	13.69	13.13	-13.13

All values are total discounted expected values in billions of 1994 U.S. dollar.

Recall that $\tau = 0$ corresponds to no experience share after the merger while $\tau = 1$ matches complete experience share. The 3 scenarios are:

- *Scenario (i):* Boeing merged with McDonnell Douglas and immediately shut down MD-11
- *Scenario (ii):* Boeing kept MD-11 after the merger.
- *Scenario (iii):* No merger

Figure 1: Interior Arrangements of a Typical Boeing 777-200 (3-Class)

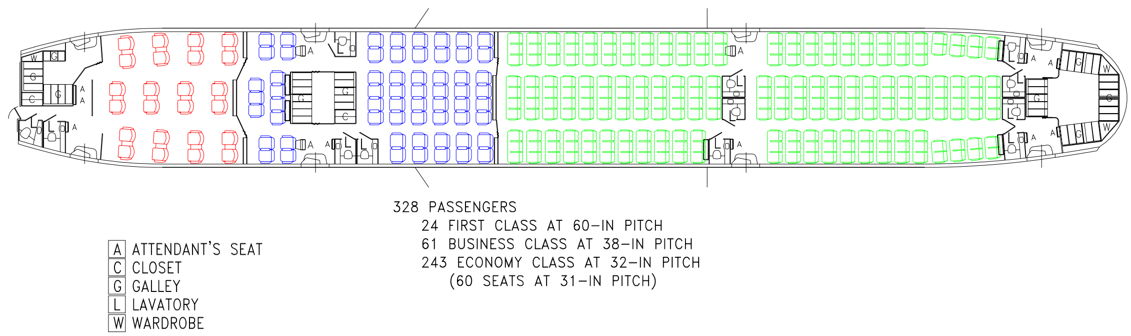


Figure 2: Seats and Range of All Wide-bodied Aircraft in Production since 1990

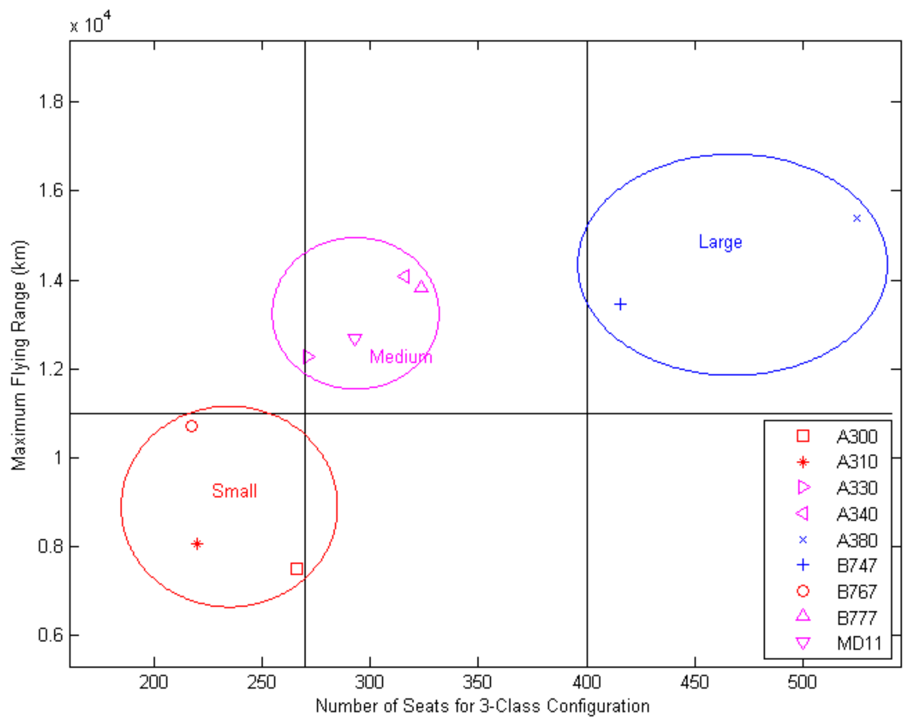


Figure 3: Distribution of *medium-wide-ratio*

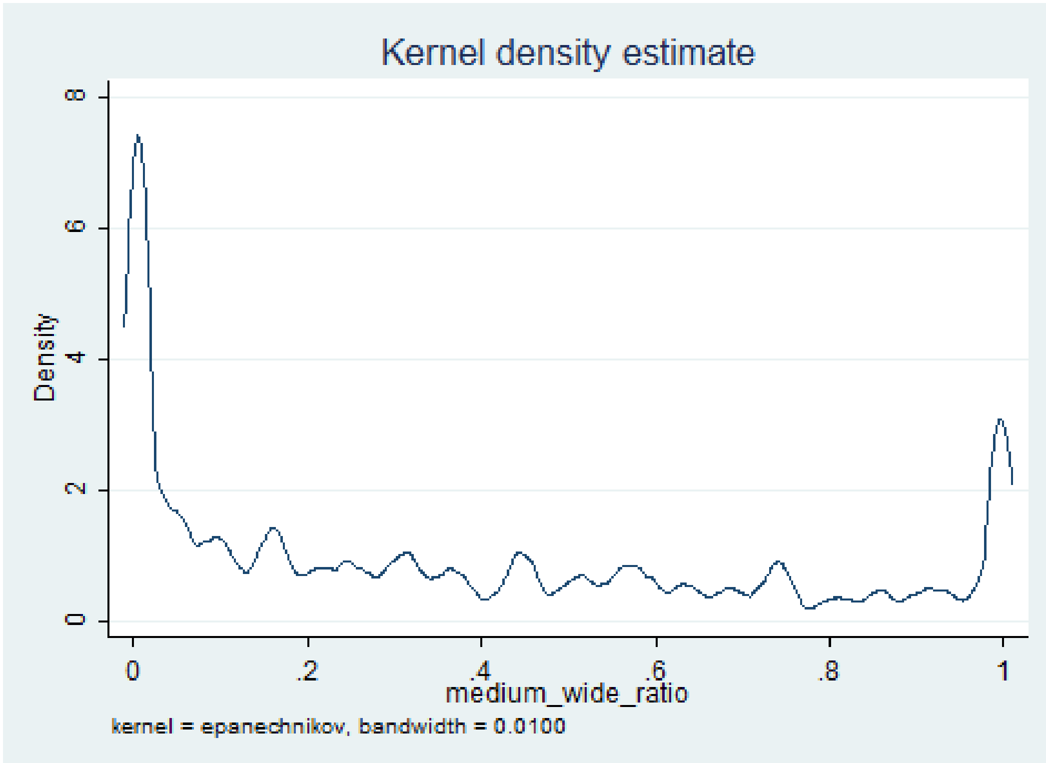


Figure 4: Correlation of Quantities and Past Accident Index

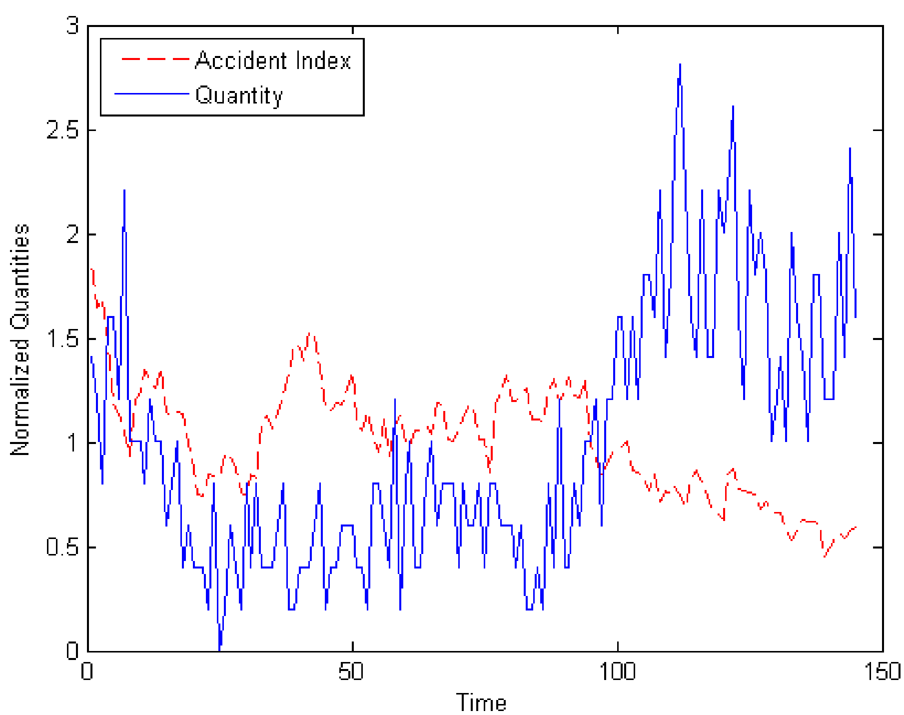


Figure 5: Fit of Labor Input of L-1011

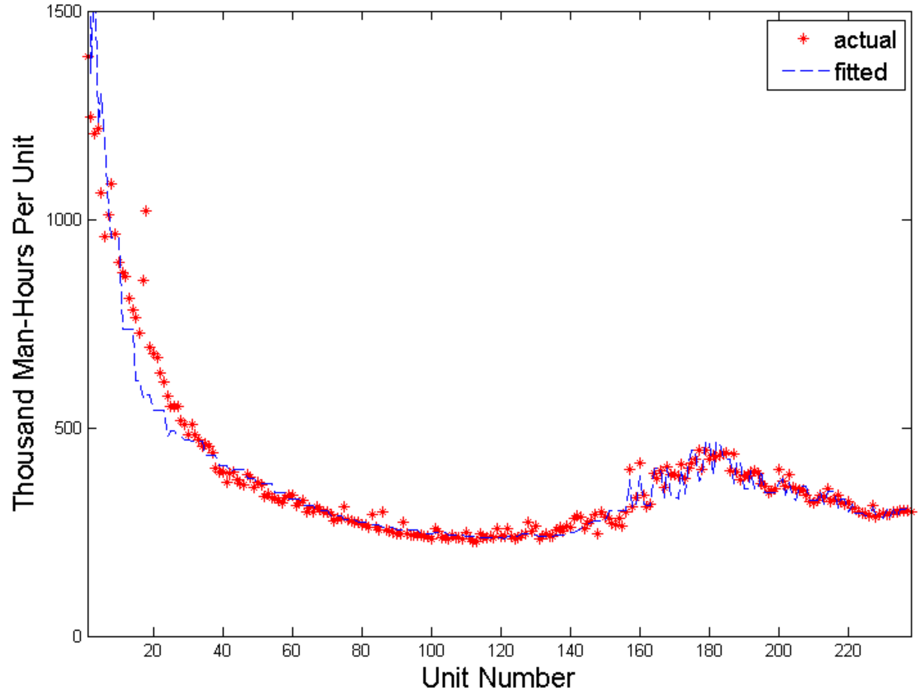


Figure 6: L-1011 Generation Impact on Experience Level

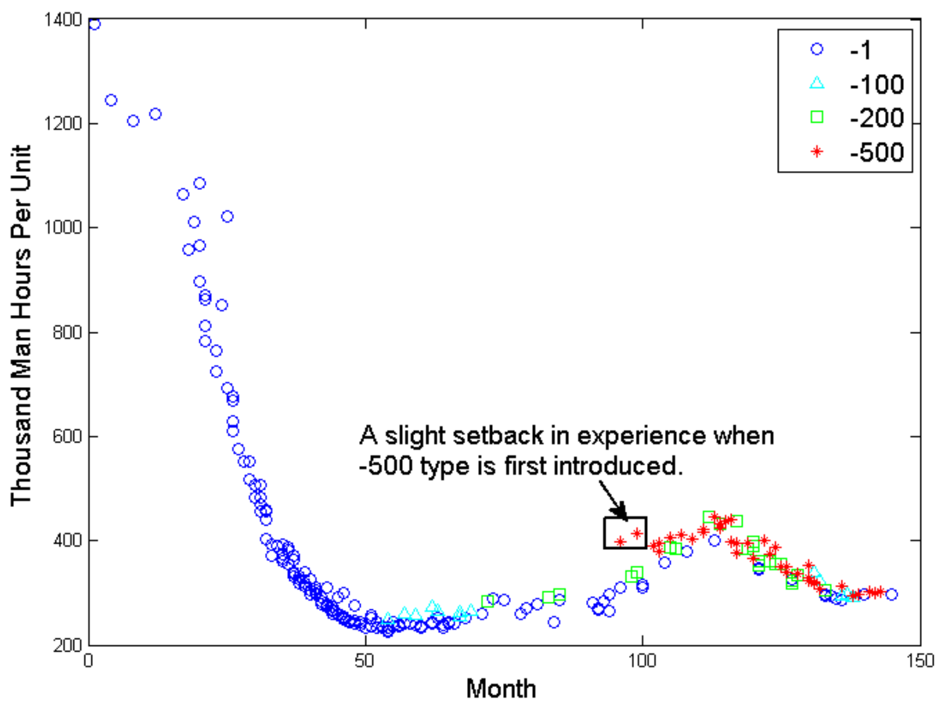


Figure 7: Distribution of Percentile Difference between Actual and Estimated Market Share Ratio

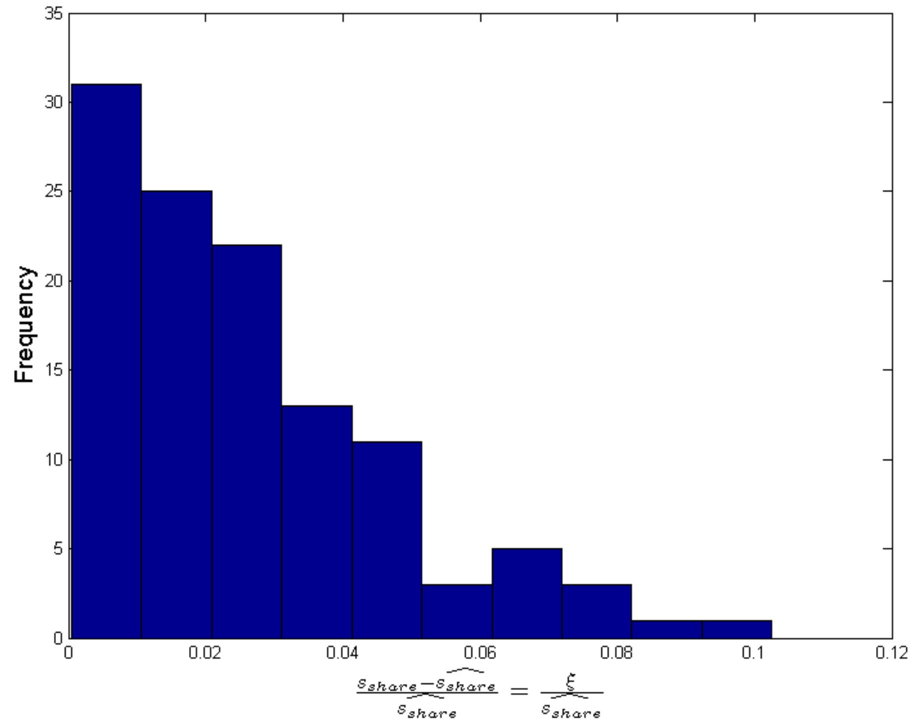


Figure 8: Distribution of Percentile Difference between Actual and Estimated Quantity

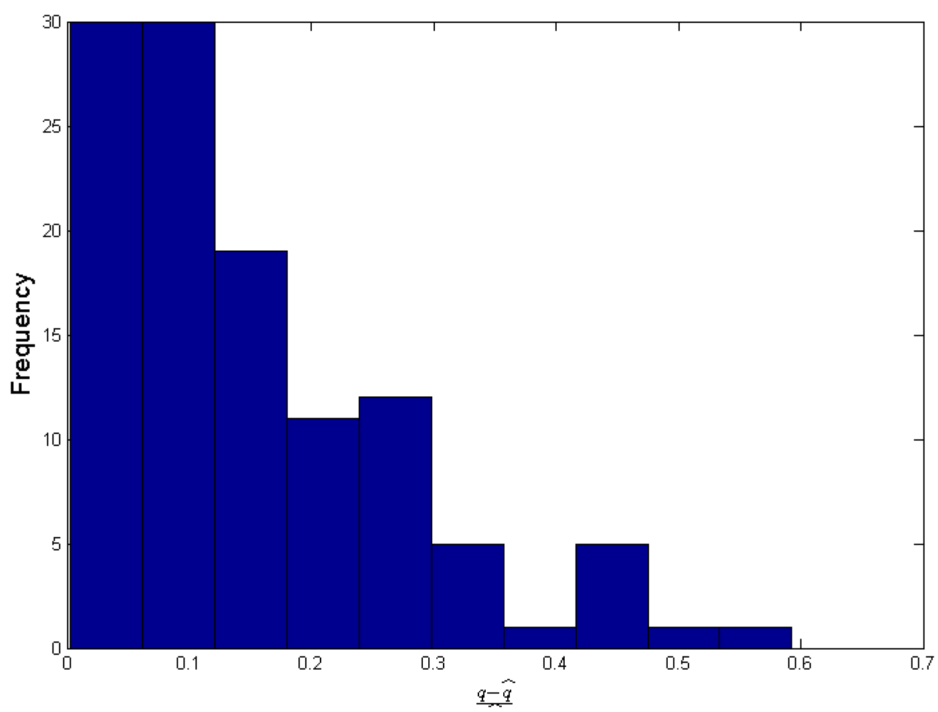


Figure 9: ξ Approximation Performance Comparison

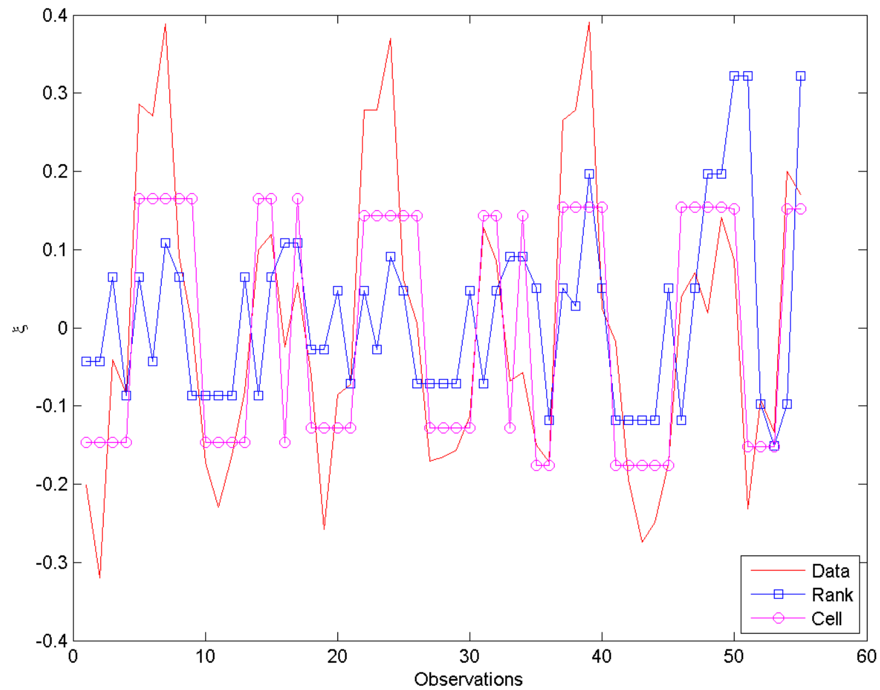


Figure 10: Demonstration of Robustness of Discretization

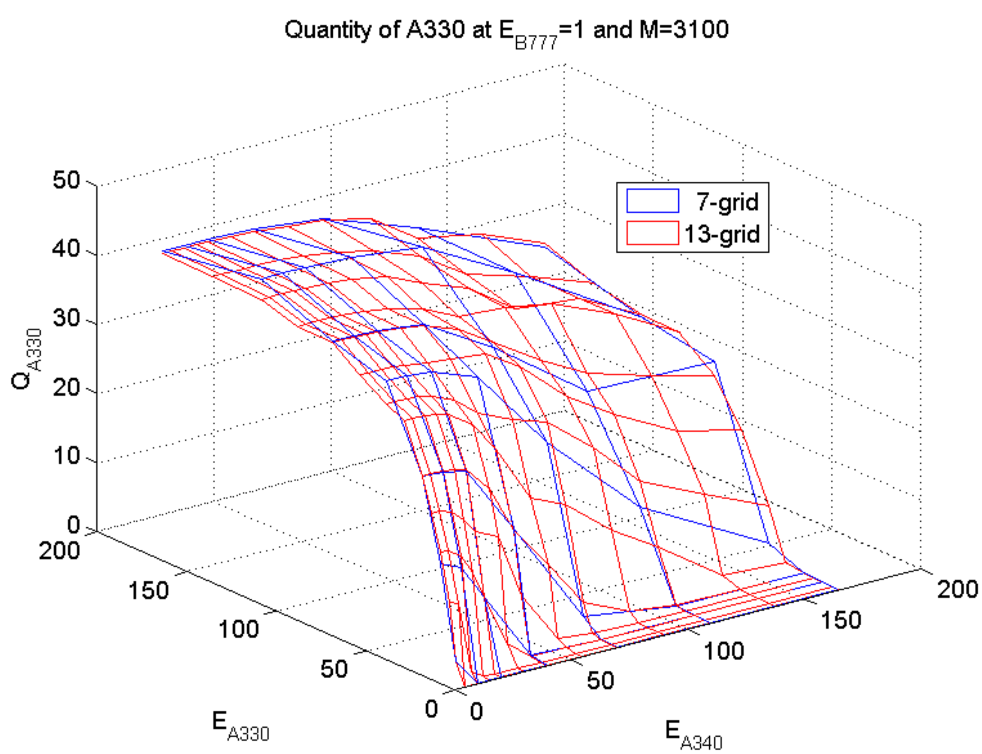


Figure 11: Comparison of ΔCS for Game A and B when τ Varies

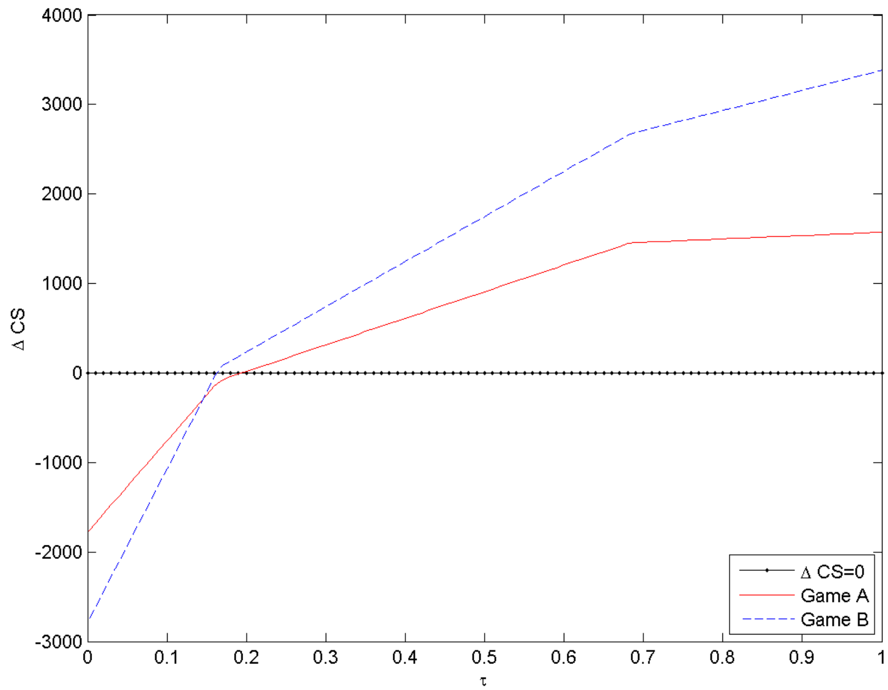


Figure 12: CS Path Comparison since 1997

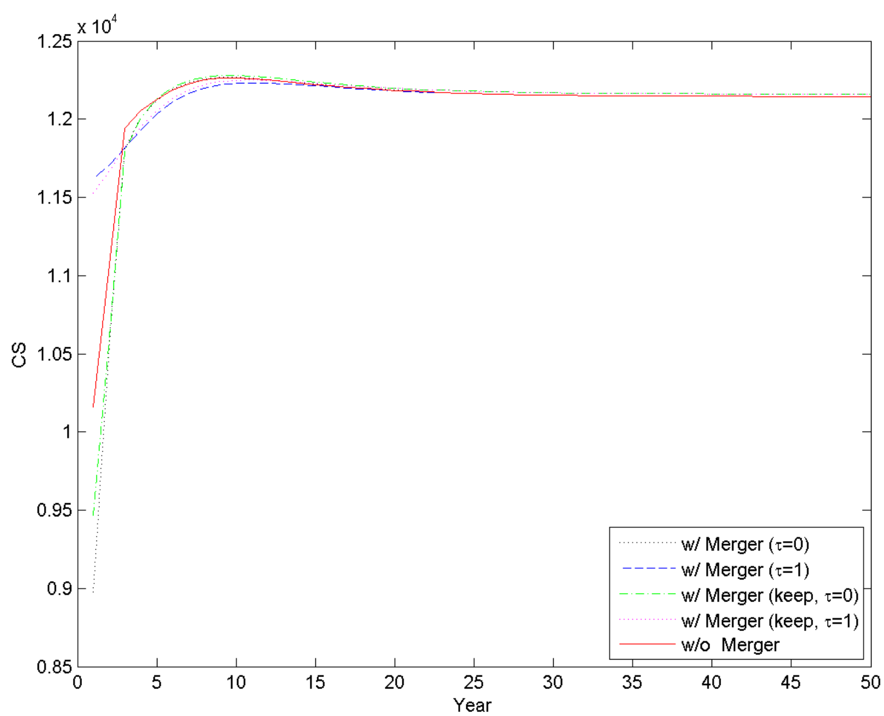


Figure 13: Quantity Path Comparison of A330 since 1997

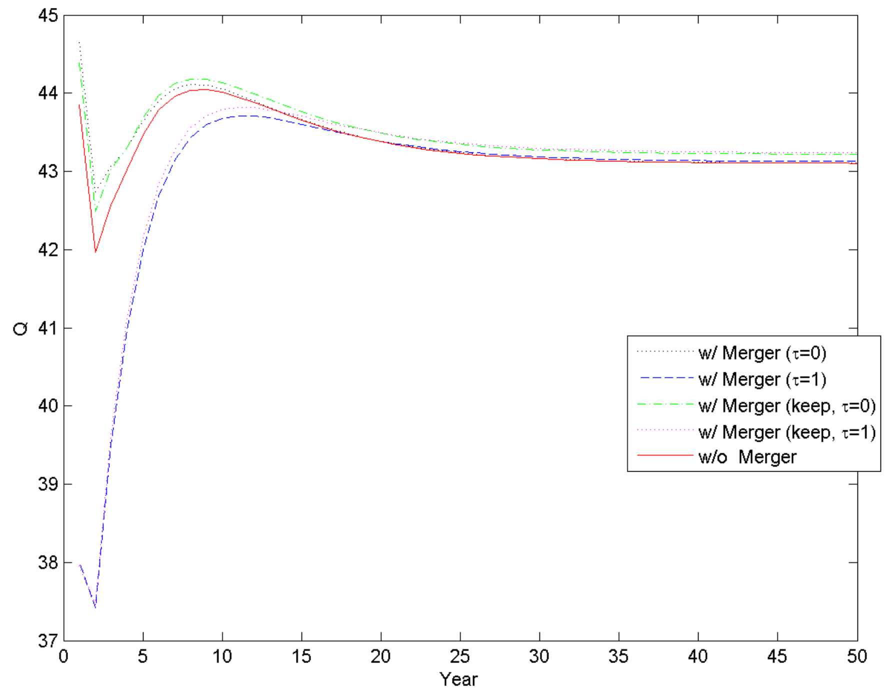


Figure 14: Quantity Path Comparison of A340 since 1997

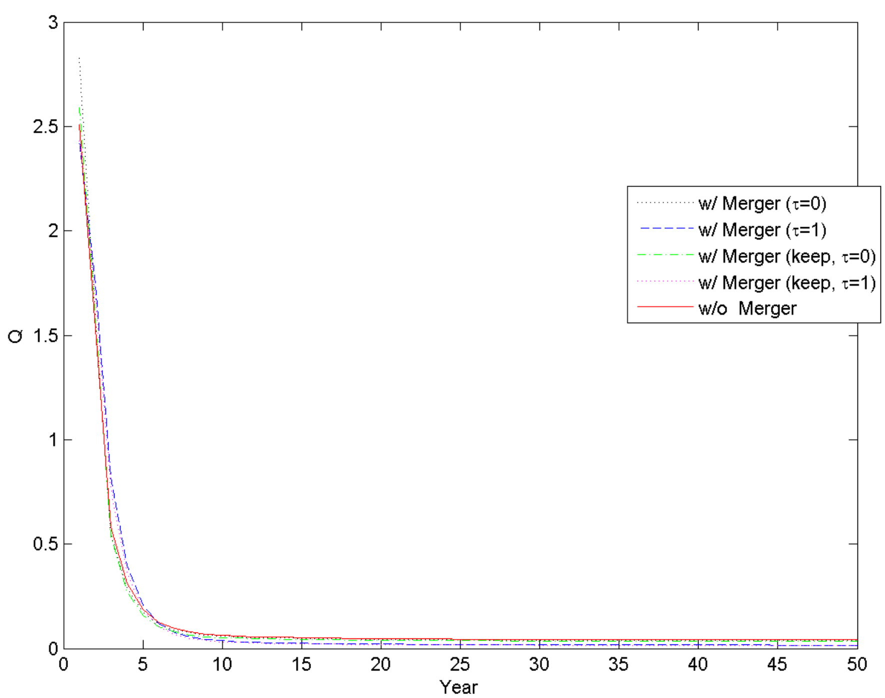


Figure 15: Quantity Path Comparison of B777 since 1997

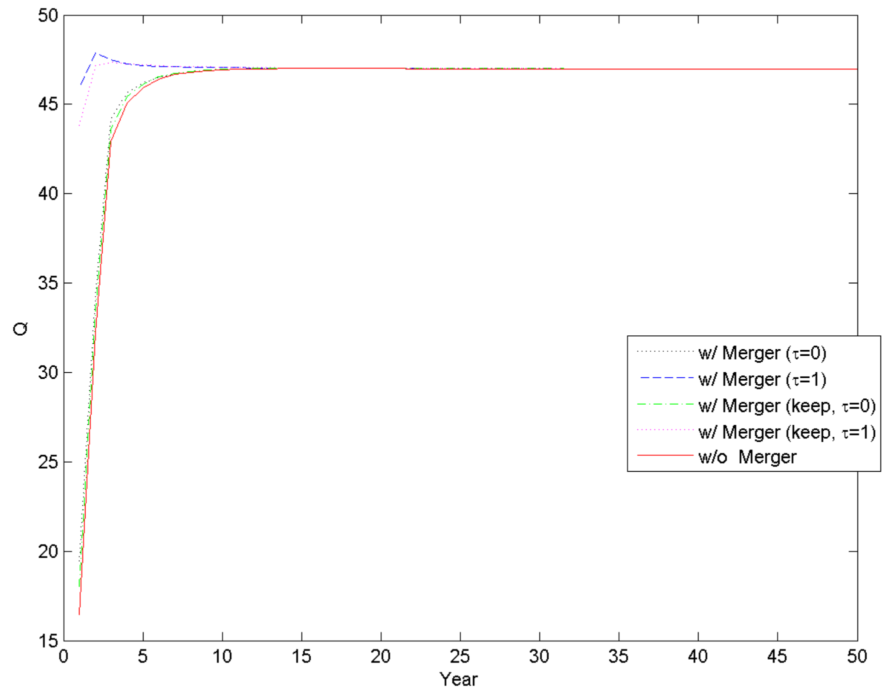


Figure 16: Quantity Path Comparison of MD-11 since 1997

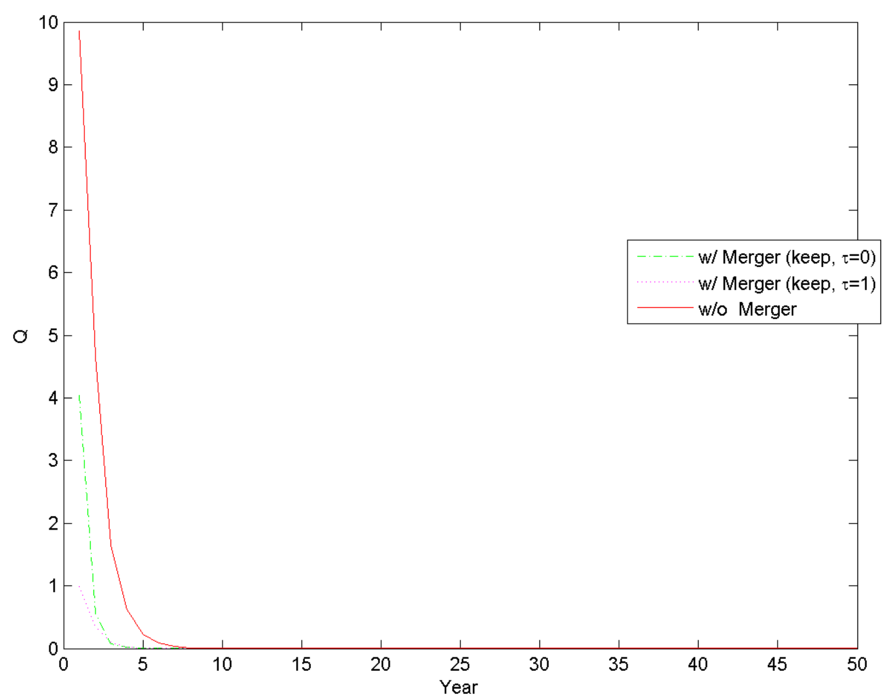


Figure 17: Experience Path Comparison of A330 since 1997

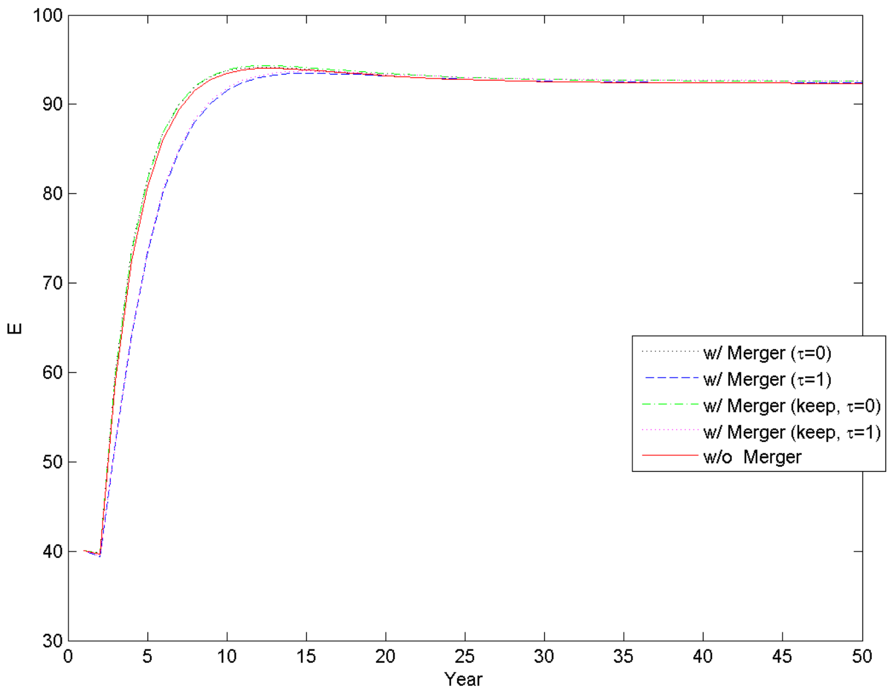


Figure 18: Experience Path Comparison of A340 since 1997

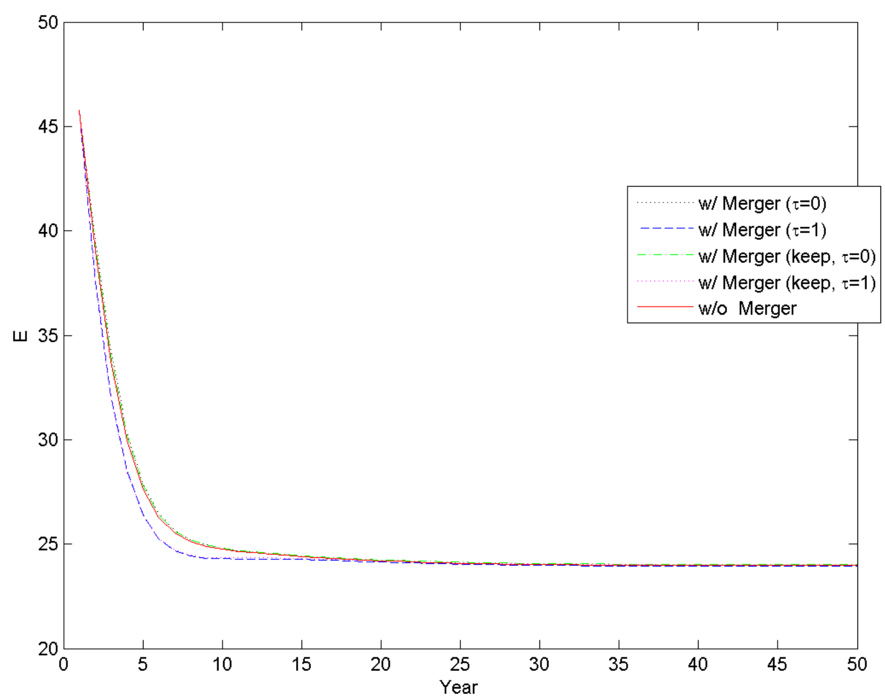


Figure 19: Experience Path Comparison of B777 since 1997

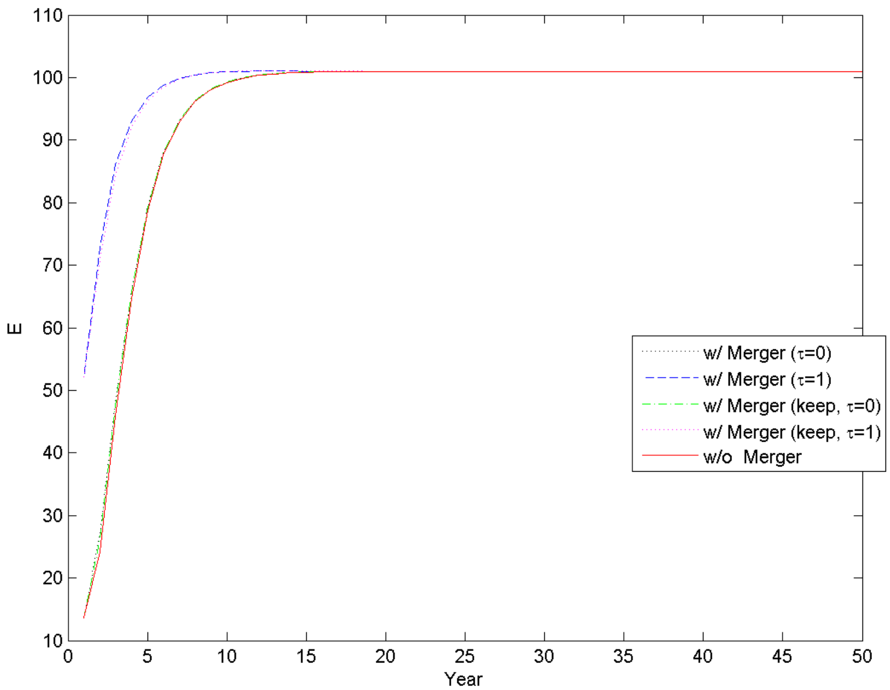


Figure 20: Experience Path Comparison of MD-11 since 1997

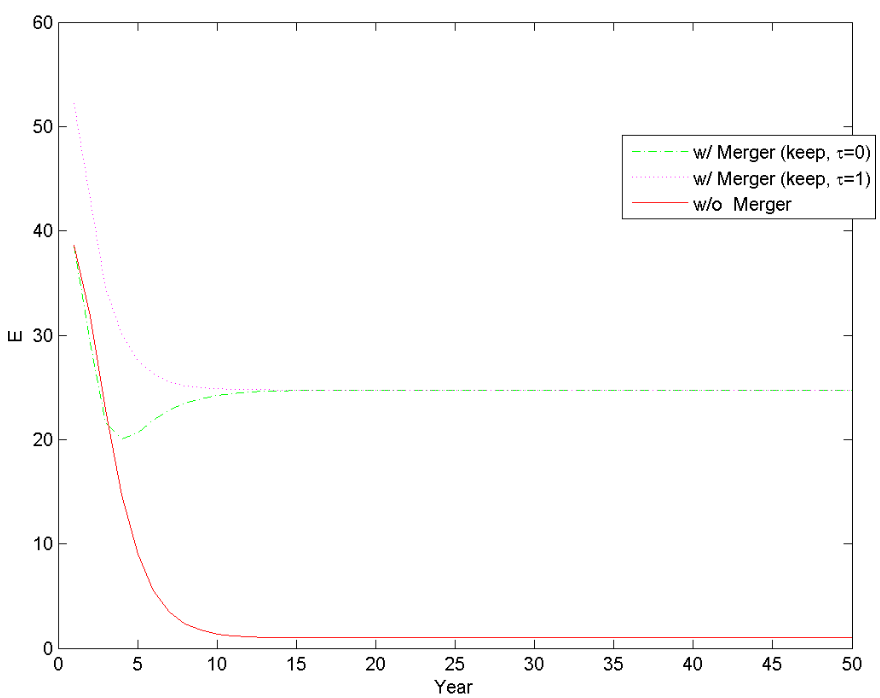


Figure 21: Paths of Expected Upgrading Prob. for A330 since 1997

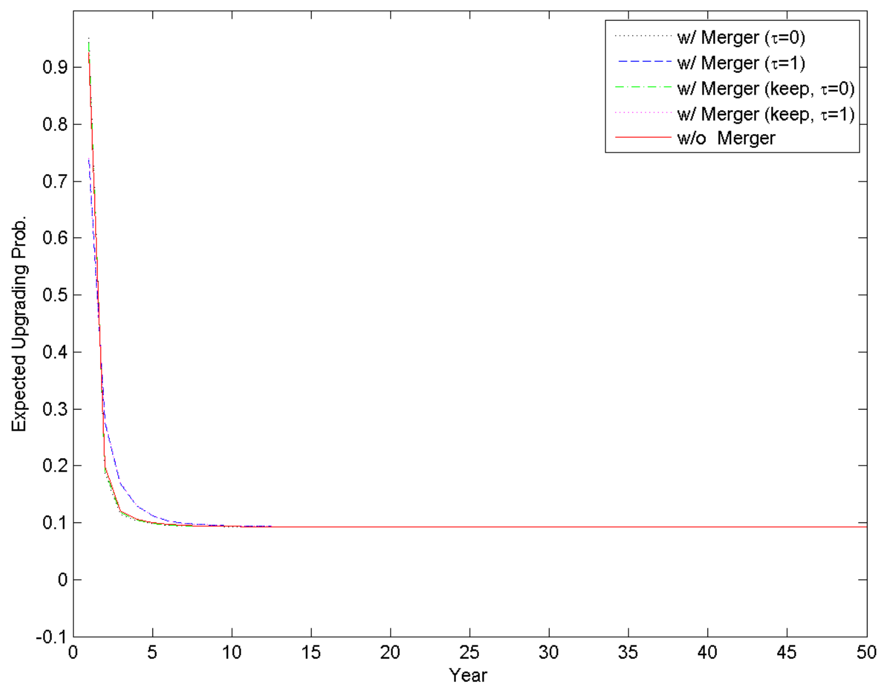


Figure 22: Paths of Expected Upgrading Prob. for A340 since 1997

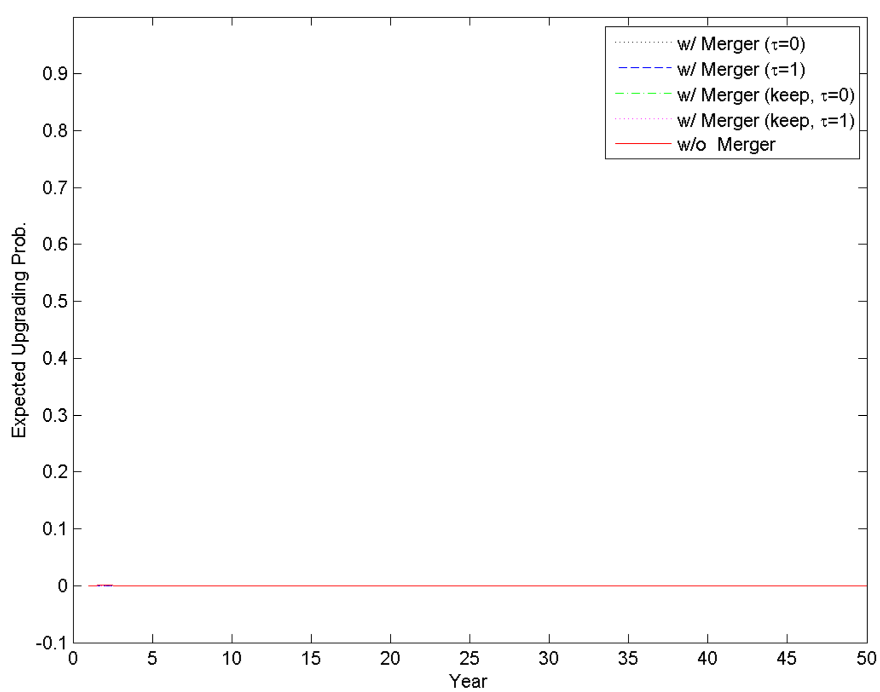


Figure 23: Paths of Expected Upgrading Prob. for B777 since 1997

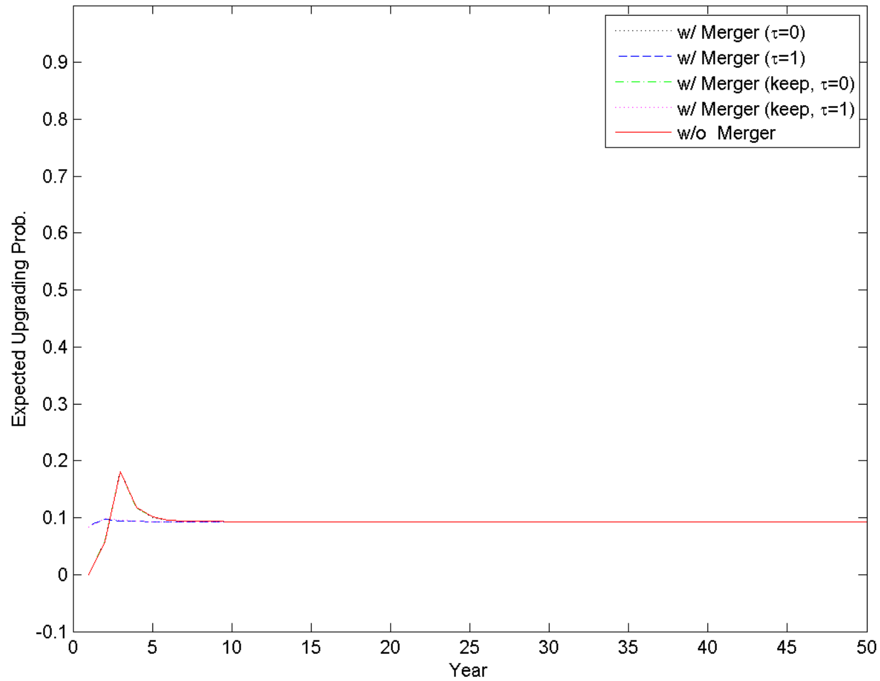


Figure 24: Paths of Expected Upgrading Prob. for MD-11 since 1997

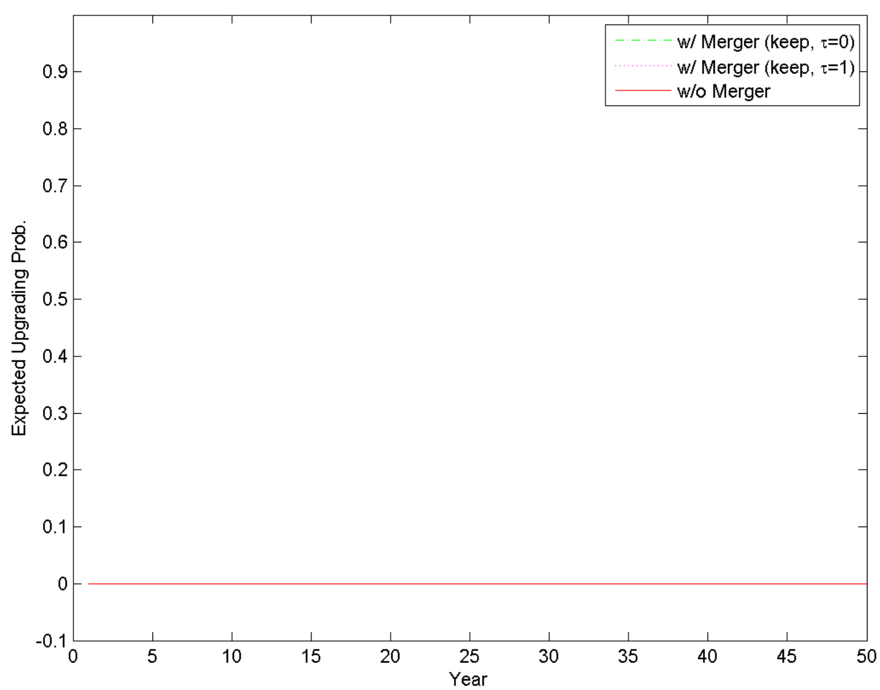


Figure 25: Comparison of ΔCS when τ Varies for Different Models

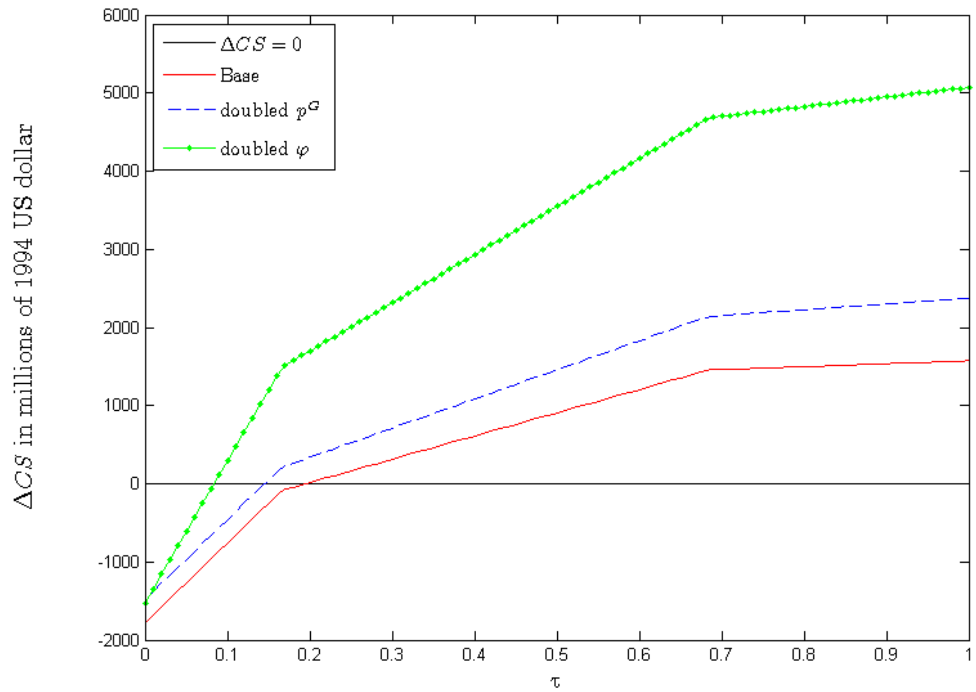


Figure 26: Merger Efficiency for different p^G and φ when $\tau = 1$

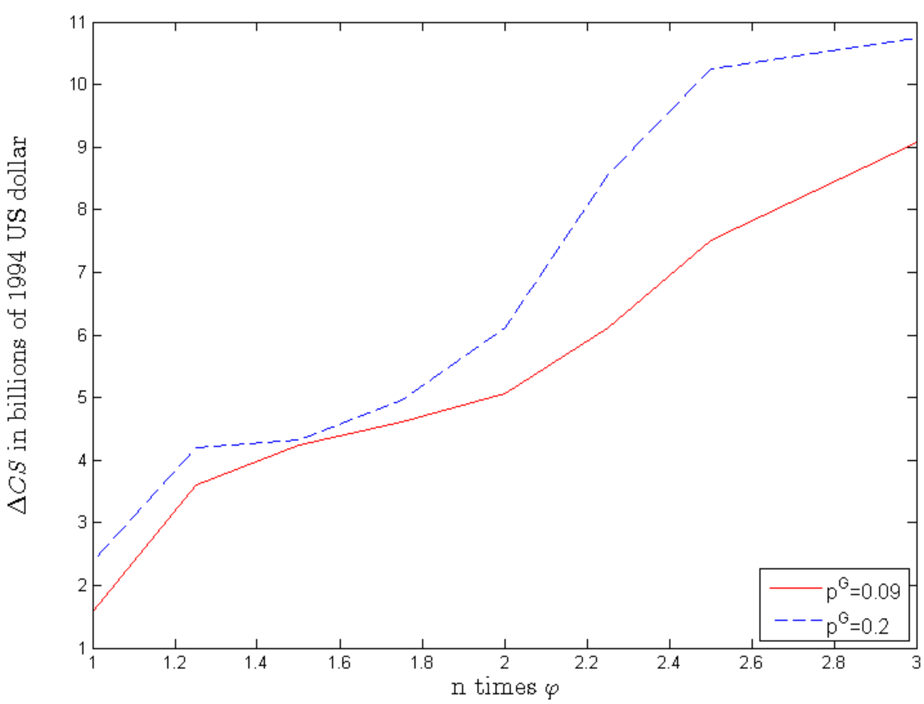


Figure 27: $EV_{00} + EV_{11} - EV_{01} - EV_{10} \geq 0$

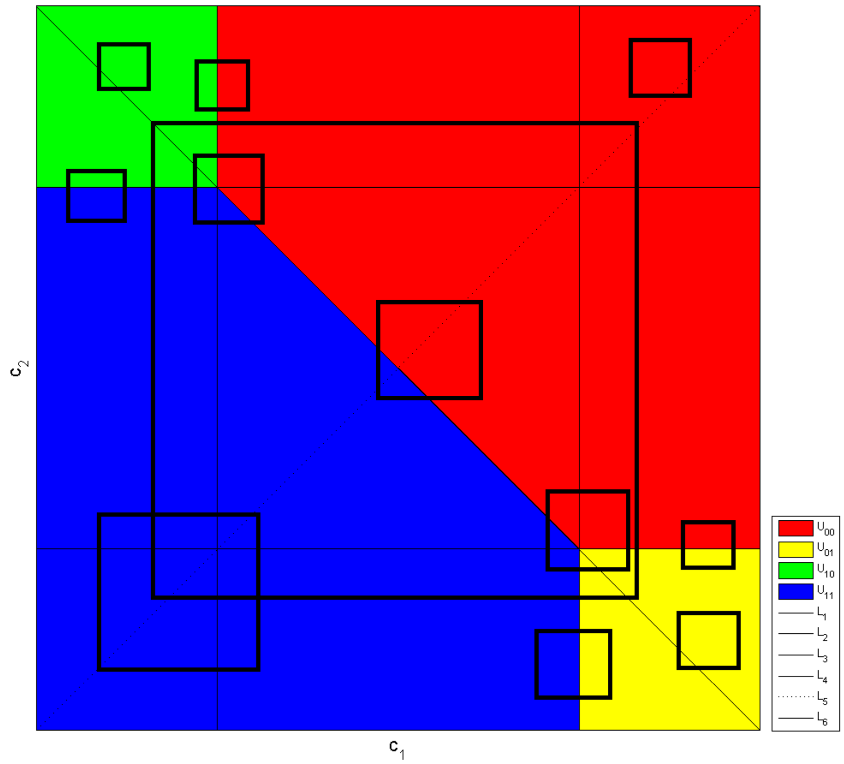
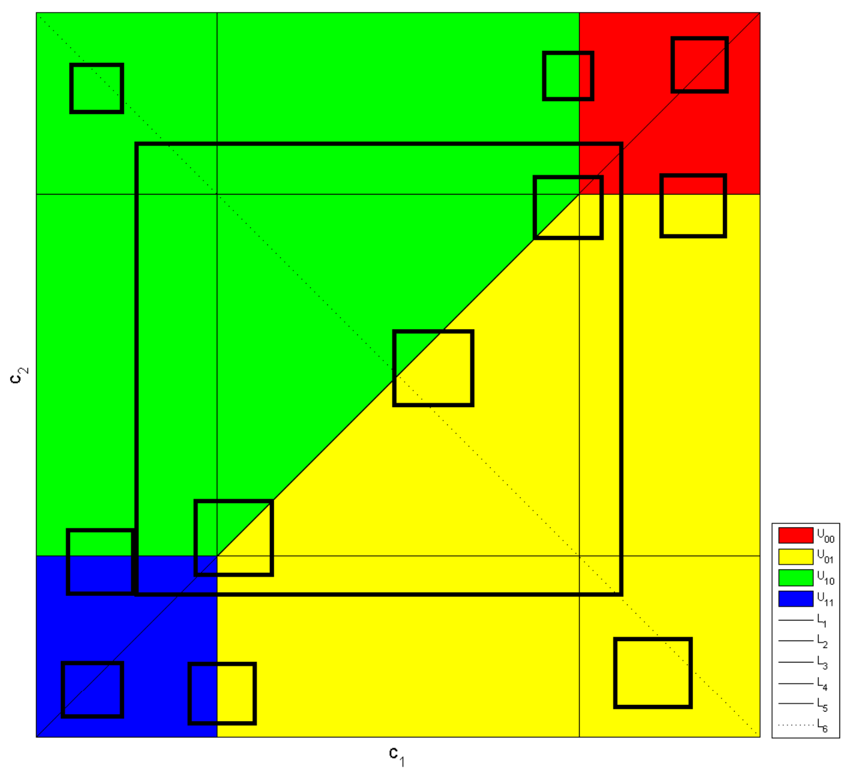


Figure 28: $EV_{00} + EV_{11} - EV_{01} - EV_{10} < 0$



3 Signaling and Tacit Collusion in an Infinitely Repeated Prisoners' Dilemma[1]

3.1 Introduction

Antitrust and competition law has long recognized that collusion comes in two varieties: explicit and tacit. Explicit collusion involves express communication among the parties regarding the collusive agreement - what outcome is to be supported and how it is to be sustained. Tacit collusion is coordination without express communication. A common form of tacit collusion is indirect communication through price signaling: A firm raises its price with the hope that other firms will interpret this move as an invitation to collude and respond by matching the price increase. As a member of the 7th Circuit Court, Judge Richard Posner articulated such a mechanism in the *High Fructose Corn Syrup* decision:

> If a firm raises price in the expectation that its competitors will do likewise, and they do, the firm's behavior can be conceptualized as the offer of a unilateral contract that the offerees accept by raising their prices.[2]

A firm raising its price in anticipation that it may be subsequently matched is taking a risk because rival firms may not respond in kind, either because they failed to properly interpret the price signal or deliberately chose not to collude. If the price rise is not matched then the firm will experience a decline in profit from a loss of demand. The prospect of such a signaling cost was well-recognized in the airlines industry where tacit collusion was implemented not with actual price increases but instead the announcement of future price increases which could be retracted (prior to any transactions taking place) in the event that rival firms did not respond with similar announcements (Borenstein, 2004). However, when such price announcements are unavailable as a signaling device, a firm must then consider the risky route of raising price without knowing how rivals will react. Of course, a firm always has the option of waiting on the hope that another firm will take the initiative of

[1] This chapter is coauthored with Joseph E. Harrington, Jr..

[2] *In Re High Fructose Corn Syrup Antitrust Litigation Appeal of A & W Bottling Inc et al*, United States Court of Appeals, Seventh Circuit, 295 F3d 651, 2002; p. 2.

raising price. The trade-off from waiting is that it avoids the possible demand loss from raising price but could delay the time until a collusive outcome is reached.

The objective of this paper is to explore the dynamics associated with the emergence of tacit collusion towards addressing the following questions: Is the likelihood of collusion declining over time? If so, does it converge to zero? If it converges to zero, does it occur asymptotically or in finite time? That is, does a sufficiently long string of failed attempts to collude result in firms, who are willing and able, becoming sufficiently discouraged that they give up trying to collude? Or is collusion assured of eventually occurring?

There are two desiderata for a theory to address these questions and shed light on the emergence of tacit collusion. First, there must be some reason for a firm to wait rather than simply set a collusive price. Second, delay must be produced so that there are some meaningful dynamics. To satisfy the first criterion, we consider an infinitely repeated two-player Prisoners' Dilemma with incomplete information. There are two player types: One type never colludes, while the other type has the capacity to collude and will surely do so once convinced its rival is also capable of colluding. Thus, the value to waiting is learning the other firm's type. As our approach will deploy the equilibrium framework, we will not be exploring the non-equilibrium process by which players settle upon a *collusive equilibrium*; players will always be playing according to some equilibrium. Tacit collusion in our setting refers to the coordination on *collusive prices* within the context of a particular equilibrium. In other words, a dynamic equilibrium process is derived that may settle on collusive prices.

The second criterion is that the theory should produce delay as an equilibrium phenomenon. Delay is required in order to have dynamics to investigate but is also desirable because it is consistent with experimental evidence and casual observation of actual markets.[3] In practice, tacit collusion takes time to develop; hence, an essential property for a theory to shed light on the emergence of collusion is that it generates delay. To satisfy this second criterion, our analysis focuses on a class of equilibria that produce delay with positive probability. A class of equilibria is considered that has two distinct phases: a learning phase

[3] The dynamics associated with the emergence of tacit collusion are difficult to document for actual markets but are well-documented for artificial markets. For the infinitely repeated Prisoners' Dilemma, experimental evidence shows that subjects may or may not cooperate and that cooperation can take time to occur; see, for example, Dal Bó (2005) and references cited therein.

and a collusion phase. In the learning phase, players are potentially signaling their types in order to initiate collusion. In the collusion phase, their types have been revealed and - if, in fact, they are collusive types - collusion subsequently occurs. This class of equilibria admits equilibria that are separating, semi-separating, and pooling. Pooling equilibria are those for which firms never collude so there is no learning phase. Separating equilibria have firms immediately reveal their types; hence, collusion occurs either without delay or not at all. While those equilibria are Pareto-efficient, the learning phase is minimal and thus they fail to produce the phenomenon that motivates the analysis. It is semi-separating equilibria that involve gradual learning and potential delay in achieving a tacitly collusive outcome. These equilibria encompass not just uncertainty about the other firm's type - is my rival willing and able to collude? - but also uncertainty about what the other firm will do - even if my rival is willing and able to collude, will it take the lead or wait for me to make the first move and raise price? This latter uncertainty arises because a collusive-type firm uses a mixed strategy that determines whether it takes the initiative by setting a high price or waits by setting a low price. Embodying both sources of relevant uncertainty makes these equilibria especially attractive in light of our objectives.[4]

To be more concrete, consider the managers of two gasoline stations located half a mile apart on the same street. Each is contemplating whether to post higher prices on its station's sign or instead deciding to "wait and see" what the other station's manager will do.[5] Is the other station also contemplating a collusive price hike but similarly holding off raising price? Or is the other station oblivious to such reasoning and has no intent of trying to tacitly collude? As time moves on without any price hikes, a station manager adjusts her beliefs as to whether the other manager is "waiting" or "oblivious" and modifies her calculus accordingly in deciding whether or not to go ahead and raise price. This is the dynamic that is captured by the equilibria characterized in this paper. Of particular

[4] As Pareto-efficiency is a common equilibrium selection device, it is important to emphasize that our selection of semi-separating equilibria is not guided by what is collectively best for firms but rather by what best matches the class of phenomena we are interested in understanding. While any delay in achieving collusive prices is Pareto-inefficent, in fact delay is a real feature to actual and experimental markets, and it is the objective of this research project to explore those dynamics.

[5] For gasoline stations in Quebec, Clark and Houde (2011) find that a small price premium (2 cents or more per liter) for a few hours can result in a significant reduction in a station's sales for the day (around 35-50%).

interest is whether, in spite of the possibility of delay, collusion will eventually occur for sure.

To summarize the main findings, the probability of collusion emerging in any period is shown to be declining over time but is always positive; at no point are beliefs sufficiently pessimistic that collusive types give up trying to collude. While always positive, the probability of collusion emerging in the current period (given it has not yet occurred) converges to zero asymptotically. Furthermore, even if both players are collusive types, the probability they *never* achieve the collusive outcome can be positive. Though collusive type players never give up trying to collude - in the sense that they always choose the collusive price with positive probability - they may never succeed in colluding. Hence, the waiting game faced by firms may not just delay collusion but prevent it from emerging altogether.

While there is a huge body of work on the theory of collusion, none of it, to our knowledge, explores the emergence of collusion through means that can reasonably be interpreted as tacit.[6] Our model does, however, share some features with the literature on reputation in that it allows private information over a player's type and the space of types includes those which are committed to a particular strategy.[7] The seminal work of Kreps et al (1982) examines cooperation in a finitely repeated Prisoners' Dilemma where an "irrational" type might be endowed with tit-for-tat, while a "rational" type optimizes unconstrained. Aumann and Sorin (1989) considered cooperation in a common interests game where a player might be endowed with a strategy with bounded recall. More recently, reputation research has considered an infinitely repeated game with commitment types with the typical research objective being to narrow down the set of equilibrium payoffs (compared to the usual Folk Theorem). When one player's type is private information, the issue is cast as whether equilibria with low payoffs for that player can be eliminated; see, for example, Cripps and Thomas (1997) and Cripps, Dekel, and Pesendorfer (2005). More recently, there has been research allowing both players to have private information; see, for example, Atakan and Ekmekci (2008). Also relevant is work on relational contracts where, in a different setting and with a different mechanism than are modelled here, players learn to cooperate more

[6] Coordination within the context of a coordination game, rather than a game of conflict, is explored in Crawford and Haller (1990).

[7] For a review of some of the research on reputation, see Mailath and Samuelson (2006).

effectively (while in our setting, they simply learn to cooperate); see Chassang (2010) and Halac (2010).

Our model considers two-sided incomplete information in the infinitely repeated setting when the commitment type is myopic. It differs in several respects from previous work on reputation. Prior research for the infinitely repeated setting has not explored the Prisoners' Dilemma but rather other stage games including games of common interests, conflicting interests,[8] and strictly conflicting interests.[9] As a result, in those settings, a player wants to mimic the commitment type, while in the PD setting, they (eventually) want to separate from the commitment type.[10] More importantly, the central issue in the reputation literature is about characterizing the set of equilibrium payoffs which, as noted above, is distinct from our objective. The task before us is not to limit the set of equilibria but rather to explore the dynamics of play for a particular class of equilibria. In our setting, a player ultimately wants to reveal it is a cooperative type but would like to do so only after the other player has done so. Thus, the issue is about the timing of building a reputation and whether that tendency to wait prevents cooperation from ever emerging. In this sense, our equilibrium has some commonality to the war of attrition characterized in Atakan and Ekmekci (2009) though they consider a different class of stage games.[11]

Though for complete information, a related mathematical structure to that explored here is Dixit and Shapiro (1985). They consider a repeated Battle of the Sexes game which can be interpreted as two players simultaneously deciding whether or not to enter a market. It is profitable for one and only one firm to enter. The stage game then has two asymmetric pure-strategy equilibria and one symmetric mixed-strategy equilibrium. In the repeated version, the dynamic equilibrium has randomization in each period with, effectively, the game terminating once there is entry. Farrell (1987) considers this structure

[8] The Stackelberg action for one player minimaxes the other player.

[9] Player 1's Stackelberg action along with player 2's best reply produces the highest stage game payoff for player 1 and the minimax payoff for player 2.

[10] On this topic, also see Mailath and Samuelson (1998).

[11] In Atakan and Ekmekci (2009), the equilibrium is equivalent to a war of attrition as each player seeks to hold out revealing it is not committed to its Stackelberg action. In their setting, the player that concedes in the war of attrition increases its current period payoff relative to not conceding, but ends up with a lower future payoff than if its rival had conceded. In our setting, the player that concedes decreases its current period payoff, relative to its rival conceding, but suffers no disadvantage in terms of its future payoff from having conceded first. In our setting, waiting occurs in order to avoid a short-run cost from conceding, while, in their setting, waiting occurs to influence the future payoff.

when players can precede their actions with messages. One can consider our equilibrium as encompassing a waiting game for which the terminal payoff (received after firms' types are common knowledge) is either the present value of the collusive payoff (when both are collusive types) or the non-collusive payoff (when one or both are non-collusive types).

After describing the model in Section 2, we define in Section 3 a class of perfect Bayesian equilibria possessing distinct learning and collusion phases. Sections 4 and 5 consider equilibria for which the learning phase is non-trivial and derives properties relating to the likelihood of collusion emerging. In Section 6, additional results are derived for some examples. Concluding remarks are provided in Section 7, and all proofs are in the appendix.

3.2 Model

Consider a two-player Prisoners' Dilemma:

Prisoners' Dilemma

Player 2

		C	D
Player 1	C	a, a	c, b
	D	b, c	d, d

where C is interpreted as the high collusive price, and D as the low competitive price. Assume $b > a > d \geq c$, and $2a \geq b + c \geq a + d$.[12] $2a \geq b + c$ is standard as it means the highest symmetric payoff has both players choosing C rather than taking turns cheating (that is, one player choosing D and the other choosing C).[13] $b + c \geq a + d$ is new and is critical to our characterization. This assumption can be re-arranged to $b - a \geq d - c$, so that the gain to playing D when the other player is expected to play C is at least as great as the gain to playing D when the other player is expected to play D. Let us show that this condition holds for both the Cournot and Bertrand oligopoly games.

Consider the symmetric Cournot quantity game with constant marginal cost c and

[12] It is typical to assume $d > c$ but we allow $d = c$. Note that we cannot have $d = c$ and $b + c = a + d$ holding simultaneously as it would then imply $b = a$, which violates the assumption that $b > a$.

[13] The condition $2a \geq b + c$ is not necessary for our results but rather is to motivate the focus on players trying to sustain (C,C) in every period.

inverse market demand for firm i of $\beta_0 - \beta_1 q_i - \beta_2 q_j$ where $\beta_0 > 0, \beta_1 \geq \beta_2 > 0$; thus, products can be differentiated. In mapping the Prisoners' Dilemma to this setting, action C corresponds to some low quantity q^l, and action D to some high quantity q^h. $b - a > d - c$ is then

$$\begin{aligned} & q^h \left[\beta_0 - \beta_1 q^h - \beta_2 q^l - c\right] - q^l \left[\beta_0 - (\beta_1 + \beta_2) q^l - c\right] \\ > \; & q^h \left[\beta_0 - (\beta_1 + \beta_2) q^h - c\right] - q^l \left[\beta_0 - \beta_1 q^l - \beta_2 q^h - c\right], \end{aligned}$$

which holds if and only if $q^h > q^l$. The Bertrand price game with homogeneous goods and constant marginal cost is, loosely speaking, the special case when $b = 2a, a > d = c = 0$. If both set the monopoly price then each earns a. Deviation from that outcome involves just undercutting the rival's price which means that the price-cost margin is approximately the same but sales are doubled so that the payoff is $2a$. Given the other firm prices at cost, pricing at cost as well yields a profit of zero (so, $d = 0$) as does pricing at the monopoly price (so, $c = 0$).[14]

Players are infinitely-lived and anticipate interacting in a Prisoners' Dilemma each period. There is perfect monitoring so the history of past actions is common knowledge. If players have a common discount factor of δ, the grim trigger strategy is a subgame perfect equilibrium if and only if:

$$\delta > \frac{b - a}{b - d}. \tag{1}$$

To capture uncertainty on the part of a player as to whether the other player is willing to cooperate, it is assumed that a player's discount factor is private information. A player can be of two possible types. A player can be type L (for "long run") which means its discount factor is δ where $\delta > \frac{b-a}{b-d}$. Or a player can be type M (for "myopic") which means its discount factor is zero (though any value less than $\frac{b-a}{b-d}$ should suffice). Hence, type M players always choose D. A necessary condition for cooperative play to emerge and persist over time is then that both players are type L.

[14] The reference to "loosely speaking" is that this interpretation requires three prices - monopoly price, just below the monopoly price, and marginal cost - while the Prisoners' Dilemma has only two actions.

3.3 A Class of Perfect Bayesian Equilibria

There are potentially many equilibria to this game and we'll focus on what we believe is a natural class in which there is a learning phase and a collusion phase.[15] The learning phase comprises those periods for which firms' types are not common knowledge and behavior depends only on beliefs over types (that is, the strategy assigns the same action for all histories that yield the same set of beliefs over types), while the collusion phase consists of periods for which firms' types are common knowledge and behavior can depend on past play in an unrestricted way. During the learning phase, players are exclusively trying to learn about the other player's type towards initiating collusion. This interpretation is made appropriate by focusing on strategies that depend only on beliefs as to the other player's type (as long as players' types are private information) and otherwise are independent of the history of play. When instead both players' types are public information, firms enter the collusion phase if they are both type L by adopting the grim trigger strategy for the remainder of the horizon. At that point, behavior depends on the history of play. Finally, there is the case when one player's type is revealed to be L and the other player's type is still private information. We will assume that both players (when they are type L) adopt the grim trigger strategy. As one player has revealed his type, the learning phase is over in which case it is natural that the player whose type has been revealed adopts a grim trigger strategy towards achieving collusion; and the other player's best response, if type L, will be to do the same.

To describe strategies during the learning phase, let α^t denote the probability that a player attaches to the other player being type L in period t. For the symmetric equilibria that we will characterize, α^t is common to both players as long as both players' types are private information. Since only type L players choose action C then if, on the equilibrium path, a player chooses C then the player must be type L. Hence, players' types are private information only as long as they have both chosen D. Given symmetric strategies (and symmetric initial beliefs), players have common beliefs regarding the other player's type, and these beliefs are common knowledge. Hence, α^t is not only the probability that player

[15] Strategies are described only when a player is type L because, when type M, a player always chooses D.

1 attaches to player 2 being type L but is also player 1's point belief as to the probability that player 2 attaches to player 1 being type L, and so forth.

The solution concept is a modification of Markov Perfect Bayesian Equilibrium (MPBE), where a strategy is Markovian only during the phase when players' types are not common knowledge. More specifically, if $\alpha^t \in (0,1)$ then a type L agent's period t play depends only on α^t and no other element of the history; a Markov strategy is then of the form, $q(\cdot) : [0,1] \to [0,1]$. As long as players' types are private information, beliefs are updated as follows. Suppose, in period t, $\alpha^t \in (0,1)$ and a type L player chooses C with probability $q^t \in (0,1)$. If a player was observed to choose D in period t then the other player updates using Bayes Rule:

$$\alpha^{t+1} = \frac{\alpha^t \left(1 - q^t\right)}{1 - \alpha^t q^t}. \tag{2}$$

Note that α^t is montonically decreasing and strictly so when $q^t \in (0,1)$. Given that a type M chooses D for sure and a type L chooses D only with probability $1-q^t$, the probability the other firm is type L is declining with the length of time for which only D has been chosen. α^1 is the common prior probability. By the usual definition, the equilibrium is not a MPBE as firms engage in a grim trigger strategy upon their types becoming common knowledge. To avoid confusion, we'll refer to the solution concept as Partial Markov Perfect Bayesian Equilibrium (PMPBE).

The class of PMPBE can be partitioned according to q^1, the probability that a type L chooses C in the first period. Initially consider a strategy profile in which $q^1 = 1$; that is, a type L player chooses C for sure. The strategy is then separating which means that the learning phase is limited to the first period. If both players choose C in period 1 then it is common knowledge both are type L and they adopt the grim trigger strategy. If, say, player 1 chooses D then player 2 assigns probability zero to player 1 being type L in which case player 2 chooses D, whether of type L or M. Thus, one or both choosing D in period 1 results in both choosing D in all ensuing periods, in which case there is no collusion.

To verify this strategy profile is an equilibrium, we need to show that choosing C for sure in period 1 is optimal and, in response to both choosing C in period 1, it is optimal for players to adopt the grim trigger strategy. Regarding period 1, $q^1 = 1$ is optimal if and only if

$$\alpha^1\left(\frac{a}{1-\delta}\right)+\left(1-\alpha^1\right)\left(c+\frac{\delta d}{1-\delta}\right)\geq\alpha^1 b+\left(1-\alpha^1\right)d+\frac{\delta d}{1-\delta}\Rightarrow \tag{3}$$

$$\alpha^1\geq\frac{(1-\delta)(d-c)}{(1-\delta)(d-c)+\delta(a-d)-(1-\delta)(b-a)} \tag{4}$$

where (4) follows from (3) assuming the denominator is positive. (If the denominator is negative then (3) does not hold.) The denominator is positive and the RHS of (4) is less than one if and only if

$$\delta(a-d)-(1-\delta)(b-a)>0\Rightarrow\delta>\frac{b-a}{b-d},$$

which we assumed in (1) to ensure that collusion is feasible under complete information. Also note that if this condition is satisfied then, in response to both choosing C in period 1, it is optimal to adopt the grim trigger strategy for the remainder of the horizon. In sum, if players are sufficiently patient (as specified in (1)) and attach sufficient probability to the other player being type L (as specified in (4)) then, when both players are type L, they will choose action C in the first period and collusion will immediately ensue. For this equilibrium, the learning phase is trivial.

Next consider a PMPBE in which $q^1=0$ so that type L players (as well as type M players) choose D in the first period. Since, by (2), $\alpha^2=\alpha^1$ then, by the Markovian assumption, $q^2=0$. By induction, $q^t=0$ for all t. This is a pooling equilibrium; it has no learning phase and firms never collude.

Finally, consider a PMPBE in which $q^1\in(0,1)$ so that a type L player assigns positive probability to both choosing C and D, so it is a semi-separating equilibrium. In that it has already been specified what happens when one or both players choose C (a player who chose C adopts the grim trigger strategy), let us explore the various possibilities when all previous play involves D having been chosen. There are three cases: i) $\exists T>1$ such that $q^t\in(0,1)$ for all $t\in\{1,...,T-1\}$ and $q^T=1$; ii) $\exists T>1$ such that $q^t\in(0,1)$ for all $t\in\{1,...,T-1\}$ and $q^T=0$; and iii) $q^t\in(0,1)$ for all t.

Case (i) has firms randomizing until period T at which time (if both have always chosen D) they choose C for sure. Let us show that such behavior cannot be part of a PMPBE.

In period $T-1$ (assuming both players chose D over periods $1, ..., T-2$), a type L firm is supposed to randomize in which case the payoffs from choosing C and D must be equal. The payoffs are

$$\begin{array}{ll} \text{Play C:} & \alpha \left[q \left(\frac{a}{1-\delta} \right) + (1-q) \left(c + \frac{\delta a}{1-\delta} \right) \right] + (1-\alpha) \left(c + \delta c + \frac{\delta^2 d}{1-\delta} \right) \\ \text{Play D:} & \alpha \left[q \left(b + \frac{\delta a}{1-\delta} \right) + (1-q) \left(d + \frac{\delta a}{1-\delta} \right) \right] + (1-\alpha) \left(d + \delta c + \frac{\delta^2 d}{1-\delta} \right) \end{array}$$

and it is clear that D yields a strictly higher payoff than C regardless of q. The reasoning is simple and standard. The only way it can be optimal to choose C is that it somehow positively influences a player's future payoff. However, when the other player is type L, a player will receive $\frac{a}{1-\delta}$ in the future whether C or D is chosen in the current period; and if the other player is type M, $c + \frac{\delta d}{1-\delta}$ is received whether C or D is chosen. Thus, D is clearly preferred. There cannot then be a PMPBE in which firms initially randomize and then adopt C for sure.

Turning to case (ii), players initially randomize and then (assuming it has been D all along) choose D for sure in period T. By our Markovian assumption, it also means choosing D in all ensuing periods. While such an equilibrium can be shown to exist by construction, this case is not very interesting for our purposes. One of our primary questions is determining whether players will eventually cooperate. By construction, this equilibrium provides a negative answer to that question by specifying that, after some series of periods in which D is chosen, players give up trying to collude and choose D for sure thereafter. What we cannot sort out with such an equilibrium is whether giving up collusion is arbitrary (each player chooses D for sure only because the other player does so) or is necessary (it is not an equilibrium for players to continue to randomize). This issue can be explored with the equilibria under case (iii).[16]

Case (iii) is when players randomize as long as D has always been chosen (and thus they are uncertain as to players' types). This is the equilibrium that will draw our attention for the remainder of the paper. It is worthy of analysis for several reasons. First, it is

[16] Case (ii) PMPBE can be shown to exist by construction using backward induction from period T. In fact, the PMPBE that we focus our attention on is the case when $T = +\infty$ and is the limit of case (ii) PMPBE as $T \to +\infty$.

useful to know whether such an equilibrium exists or instead equilibria must be of the form in case (ii) which would imply that beliefs must eventually become sufficiently pessimistic that attempts at collusion stop. Second, if these equilibria do exist - so players keep on trying to collude in the sense of choosing C with positive probability - there is the question of whether it implies that collusion will eventually occur for sure. Even if the probability of choosing C declines over time, whether collusion is ensured depends on the speed of that decline. Third, the primary focus of the paper is on the learning phase, which makes this equilibrium attractive because learning is not arbitrarily assumed to terminate in some period (as with case (ii)). Instead, firms randomize as long as it is optimal to do so which continues to provide the opportunity to learn a rival's type. Fourth, in contrast to the Pareto-efficient separating equilibrium in which collusive types collude for sure in the first period, this equilibrium is able to produce delay with positive probability which means we can explore the dynamics associated with the emergence of tacit collusion. Though firms may prefer to enact collusion without delay, market and experimental evidence show that they often do not. An equilibrium with delay may then be able to deliver some insight regarding the emergence of collusion.

3.4 Equilibrium Properties

In this section we explore some properties of a Partial Markov Perfect Bayesian Equilibrium. Recall that a PMPBE is partly described by: if $\alpha^t \in (0,1)$ then a type L agent's period t play depends only on α^t and no other element of the history, so it is of the form, $q(\cdot) : [0,1] \to [0,1]$. The particular class of PMPBE we will explore are defined by the following properties. In period 1, choose C with probability $q\left(\alpha^1\right) \in (0,1)$. In period $t \geq 2$, if (D,D) in all previous periods then choose C with probability $q\left(\alpha^t\right) \in (0,1)$; and if (D,D) in periods $1, ..., t-2$ and not (D,D) in period $t-1$ then choose C and adopt the grim trigger strategy. Recall that, as long as both players chose D, α^t evolves according to (2). Equilibrium conditions are of three types. First, conditions to ensure randomization is optimal when (D,D) has always been played. Second, given both players chose D up to the preceding period and then one player chose C and the other chose D in the preceding period, it is optimal for the player who chose C to do so again in the current period (it being the initial move for the grim trigger strategy). Third, in response to the history just described, it is optimal for the player who chose D to choose C (again, it being the initial move for the grim trigger strategy). The last scenario just requires optimality of the grim trigger strategy given the other player is type L and chooses the grim trigger strategy, which is satisfied if and only if (1) holds. The second case is distinct in that player 1 remains uncertain as to the other player's type. After dealing with the first set of conditions, we'll examine the second condition.[17]

Before tackling these conditions, a comment is in order. In deriving equilibrium conditions, a player will go through the thought experiment of deviating from $q(\cdot)$. Note, however, that this does not upset the specification of common beliefs. Suppose player 1 deviates in period t by not choosing C with probability $q\left(\alpha^t\right)$. As each player expects the other to have chosen C with probability $q\left(\alpha^t\right)$, each player assigns probability $\frac{\alpha^t\left(1-q\left(\alpha^t\right)\right)}{1-\alpha^t q\left(\alpha^t\right)}$ to the other player being type L. While player 1 knows that player 2's beliefs about player 1's type are incorrect, that is irrelevant as all player 1 cares about is player 2's type and player 2's beliefs, both of which are summarized by $\frac{\alpha^t\left(1-q\left(\alpha^t\right)\right)}{1-\alpha^t q\left(\alpha^t\right)}$. Thus, $\frac{\alpha^t\left(1-q\left(\alpha^t\right)\right)}{1-\alpha^t q\left(\alpha^t\right)}$ remains the relevant state variable, even if a player deviates from equilibrium play.

Suppose both players' types are private information, so either it is period 1 or it is some future period but both players have thus far only chosen D. A player's expected payoff from choosing C is

$$W^C(\alpha) \equiv \alpha\left[q\left(\frac{a}{1-\delta}\right)+(1-q)\left(c+\frac{\delta a}{1-\delta}\right)\right]+(1-\alpha)\left(c+\delta c+\frac{\delta^2 d}{1-\delta}\right).$$

With probability α, the other player is type L and chooses C with probability q which results in cooperative payoff a being earned in the current and future periods; and chooses D with probability $1-q$ so that payoff c is earned in the current period and the cooperative payoff

[17] As shown by, for example, Bhaskar (1998) and Bhaskar, Mailath, and Morris (2008), mixed strategy equilibria for an infinitely repeated game need not be purifiable, which, if that is the case, removes an important motivation for mixed strategy equilibria. The loss of purification is due to the loss of local uniqueness of Nash equilibrium. For example, Bhaskar (2000) derived a continuum of mixed strategy Nash equilibria for a repeated game, none of which were the limit of pure strategy equilibria of a perturbed game. This concern about purification could well provide a rationale for our focus on PMPBE. With PMPBE, randomization only occurs when strategies condition on players' (common) belief over the other player's type. While we have not proven local uniqueness of such equilibria, it would be surprising if that was not the case.

thereafter. Note that, regardless of the other player's action, if the other player is type L as well then both players adopt the grim trigger strategy thereafter so a is earned in the future. With probability $1-\alpha$, the other player is type M so that player chooses D which results in a payoff of c in the current and subsequent period (as C is chosen in the next period as well on the hope that collusion will have been initiated) and the non-collusive payoff d thereafter. Simplifying this expression,

$$W^C(\alpha) = \alpha q\,(a-c) + (1+\delta)\,c + \delta\alpha\left[\frac{(a-d)}{1-\delta} + (d-c)\right] + \frac{\delta^2 d}{1-\delta}. \tag{5}$$

The expected payoff from choosing D is

$$\begin{aligned} W^D(\alpha) \;\; &\equiv \;\; \alpha\left[q\left(b+\frac{\delta a}{1-\delta}\right) + (1-q)\left(d+\delta V\left(\frac{\alpha\,(1-q)}{1-\alpha q}\right)\right)\right] \\ &\quad + (1-\alpha)\left(d+\delta V\left(\frac{\alpha\,(1-q)}{1-\alpha q}\right)\right), \end{aligned} \tag{6}$$

where $V:[0,1]\to\Re$ is the PMPBE value function for a type L player. (6) can be simplified to

$$W^D(\alpha) = \alpha q\left(b+\frac{\delta a}{1-\delta}\right) + (1-\alpha q)\left(d+\delta V\left(\frac{\alpha\,(1-q)}{1-\alpha q}\right)\right). \tag{7}$$

If, in equilibrium, $q\in(0,1)$ then the expressions in (5) and (7) must be the same:

$$\begin{aligned} &\alpha q\,(a-c) + (1+\delta)\,c + \delta\alpha\left[\frac{(a-d)}{1-\delta} + (d-c)\right] + \frac{\delta^2 d}{1-\delta} \\ = \;\; &\alpha q\left(b+\frac{\delta a}{1-\delta}\right) + (1-\alpha q)\left(d+\delta V\left(\frac{\alpha\,(1-q)}{1-\alpha q}\right)\right). \end{aligned}$$

Re-arranging gives us:

$$\alpha q = \frac{\delta\left[\frac{a}{1-\delta} - V\left(\frac{\alpha(1-q)}{1-\alpha q}\right)\right] - (1-\alpha)\,\frac{\delta(a-d)}{1-\delta} - \left[1+\delta\,(1-\alpha)\right](d-c)}{\delta\left(\frac{a}{1-\delta} - V\left(\frac{\alpha(1-q)}{1-\alpha q}\right)\right) + (b-a) - (d-c)}. \tag{8}$$

(8) will need to hold for type L firms to optimally randomize when the history is composed only of having played D.

Next consider the situation in which, prior to the previous period, (D,D) had always been played so that both players' types were private information, and, in the previous period, one player chose C and the other chose D. If both are type L, they adopt the grim trigger strategy. For the player who knows the other player's type (and thus knows he'll choose C), choosing C is optimal if and only if (1) holds. Now consider the player whose type has been revealed and remains uncertain as to the other player's type. If that player assigns probability α to the other player being type L then he prefers to choose C if and only if

$$\alpha\left(\frac{a}{1-\delta}\right)+(1-\alpha)\left(c+\frac{\delta d}{1-\delta}\right)\geq\alpha\left(b+\frac{\delta d}{1-\delta}\right)+(1-\alpha)\left(\frac{d}{1-\delta}\right)$$

which is equivalent to[18]

$$\alpha\geq\frac{(1-\delta)(d-c)}{(1-\delta)(d-c)+[\delta(a-d)-(1-\delta)(b-a)]}\equiv\alpha^{*}. \tag{9}$$

(9) will need to hold for type L firms to optimally adopt the grim trigger strategy when one player chose C in the preceding period.

Theorem 1 states that there is a symmetric PMPBE in which, as long as players' types are private information, a type L player randomizes between playing C and D when $\alpha>\underline{\alpha}$, and chooses D for sure when $\alpha\leq\underline{\alpha}$ (where $\underline{\alpha}$ is defined in Theorem 1). When a player randomizes, C is chosen with probability $q(\alpha)$, as defined in (8). Proofs are in the appendix.

Theorem 1 *There exists $\widehat{\delta}\in(0,1)$ such that if $\delta>\widehat{\delta}$ then there is a symmetric Partial Markov Perfect Bayesian Equilibrium $q(\cdot)$ such that*

$$q(\alpha)\begin{cases}=0 & \text{if } \alpha\in(0,\underline{\alpha}]\\ \in(0,1) & \text{if } \alpha\in(\underline{\alpha},1]\end{cases}$$

$$V(\alpha)\begin{cases}=\frac{d}{1-\delta} & \text{if } \alpha\in(0,\underline{\alpha}]\\ \in\left(\frac{d}{1-\delta},\frac{a}{1-\delta}\right) & \text{if } \alpha\in(\underline{\alpha},1]\end{cases}$$

[18] Note that $\delta(a-d)-(1-\delta)(b-a)>0$ as it is equivalent to $\delta>\frac{b-a}{b-d}$, which holds by assumption. Hence, $\alpha^{*}\in(0,1)$.

where

$$\underline{\alpha} \equiv \frac{\left(1-\delta^{2}\right)(d-c)}{\delta\left[(1-\delta)(a-c)+\delta(a-d)\right]} \in [0,1),$$

and $\lim_{\alpha\to 1} q(\alpha) < 1$.

Let us briefly review the key elements of the proof of Theorem 1. As stated in Theorem 1, if $\alpha < \underline{\alpha}$ then firms choose D. To ensure the optimality of that behavior, the highest value for α is found such that a firm prefers D regardless of the value for q of the other firm (with the binding case being $q = 0$). This condition delivers $\underline{\alpha}$. Theorem 1 also states that, when $\alpha > \underline{\alpha}$, a firm is content to randomize between C and D given the history is composed of firms having only chosen D. It is shown that if $\alpha > \underline{\alpha}$ then, if its rival chooses D for sure ($q = 0$), a firm strictly prefers C ($q = 1$) because there'll be no chance of cooperation otherwise; thus, $q = 0$ is not part of a symmetric equilibrium. If its rival chooses C for sure ($q = 1$), a firm strictly prefers D ($q = 0$) because it does not need to choose C in order for cooperation to emerge; thus, $q = 1$ is not part of a symmetric equilibrium either. From these results and the continuity of payoffs, it follows that there is a value of $q \in (0,1)$ such that the expected payoffs from C and D are equalized. Finally, recall that when a player has randomized and chosen C when the other firm simultaneously chose D, the player that chose C needs to find it optimal to choose C in the next period for sure. It was previously derived in (9) that the prescribed behavior of following C with C is optimal if and only if $\alpha \geq \alpha^{*}$. It is shown in the proof that if δ is sufficiently close to one then $\underline{\alpha} \geq \alpha^{*}$ which ensures that a firm that takes the initiative by first choosing C (which implies $\alpha > \underline{\alpha}$) will optimally choose C again on the hope that its rival will reciprocate by choosing C.

The next result concerns the evolution of beliefs and behavior in response to a failure to cooperate, by which we mean both players have thus far always chosen D. Recall that if a player assigns probability α to the other player being type L then, after observing the other player choose D, the updated probability is $\frac{\alpha(1-q(\alpha))}{1-\alpha q(\alpha)}$ where $q(\alpha)$ is the equilibrium probability that a type L player chooses C given beliefs α. Further recall, from Theorem 1, that if $\alpha > \underline{\alpha}$ then $q(\alpha) > 0$.

Theorem 2 *If $q(\cdot)$ is a symmetric Partial Markov Perfect Bayesian Equilibrium as described in Theorem 1 then: i) if $\alpha^{1} > \underline{\alpha}$ then $\alpha^{t} > \alpha^{t+1}$ and $q\left(\alpha^{t}\right) > 0$ for all t and*

$lim_{t\to\infty}\alpha^t = \underline{\alpha}$; ii) if $\underline{\alpha} > 0$ then $lim_{\alpha\downarrow\underline{\alpha}} q(\alpha) = 0$; and iii) $\lim_{t\to\infty}\alpha^t q\left(\alpha^t\right) = 0$.

Recall that $\underline{\alpha}$ has the property that if $\alpha^t < \underline{\alpha}$ then choosing D is optimal because the likelihood that its rival is type L - and has the capacity to collude - is sufficiently low. Hence, if $\alpha^1 < \underline{\alpha}$ then type L firms will choose D in the first period (and, by stationarity, thereafter) and collusion never has a chance to emerge. However, if instead $\alpha^1 > \underline{\alpha}$ then, by Theorem 2, $\alpha^t > \underline{\alpha}$ for all t which then implies $q\left(\alpha^t\right) > 0$ for all t. Therefore, no matter how long players have failed to cooperate, a type L player will continue to try to initiate cooperation (in the sense of assigning positive probability of choosing C). In other words, beliefs never become so pessimistic about the other player's willingness to cooperate that a player prefers to abandon any prospects of cooperation by playing D for sure. When $\underline{\alpha} > 0$, it is also the case that the probability of a player initiating cooperation converges to zero over time in response to the probability that the other player is type L converging to $\underline{\alpha}$ after a history of failed cooperation. Note that the probability of a type L player playing C must converge to zero as the probability of a player being type L approaches $\underline{\alpha}\,(> 0)$ from above. If $q(\alpha)$ was instead bounded above zero then a sufficiently long sequence of playing D would have to result in a sufficiently small probability of the player being type L, which would contradict this probability being bounded below by $\underline{\alpha}$ (at least when $\underline{\alpha} > 0$). Finally, conditional on cooperation not yet having emerged, the probability assigned to a player initiating cooperation is $\alpha^t q\left(\alpha^t\right)$ in which case the probability that cooperation emerges out of period t is $1 - \left(1 - \alpha^t q\left(\alpha^t\right)\right)^2$. While this value is always positive - so collusion is always a possibility - it converges to zero in response to an ever-increasing sequence of failed attempts at collusion, in which case collusion eventually becomes very unlikely to emerge. Whether collusion emerges for sure is explored for the class of PMPBE examined in the next section.

3.5 Affine Equilibria

For the class of PMPBE described in Theorem 1, let us examine those for which the value function is affine in α (when players' types are private information). The appeal to affine PMPBE is their tractability in that they have closed-form solutions, which will allow additional properties to be derived about the dynamics. In particular, the question of whether collusion is delayed but inevitable can be addressed.

Definition 3 *An affine Partial Markov Perfect Bayesian Equilibrium is a PMPBE (as described in Theorem 1) in which the value function is affine in* α *for* $\alpha \in [\underline{\alpha}, 1]$.

Theorem 4 *There exists* $\widehat{\delta} \in (0,1)$ *such that if* $\delta > \widehat{\delta}$ *then there exists a unique affine Partial Markov Perfect Bayesian Equilibrium. The value function is*

$$V(\alpha) = \begin{cases} \frac{d}{1-\delta} & \text{if } \alpha \in [0, \underline{\alpha}] \\ x + y\alpha & \text{if } \alpha \in [\underline{\alpha}, 1] \end{cases} \tag{10}$$

where (x, y) *is the unique solution to:*

$$x + y\frac{\left(1-\delta^2\right)(d-c)}{\delta\left[(1-\delta)(a-c)+\delta(a-d)\right]} = \frac{d}{1-\delta} \tag{11}$$

$$x + y = \frac{2a\delta + (1-\delta)\left[(b-a)-(d-c)-\sqrt{\Omega}\right]}{2\delta(1-\delta)}. \tag{12}$$

and

$$\Omega \equiv \left[(b-a)-(d-c)\right]^2 + 4\delta(a-c)(b-a). \tag{13}$$

Furthermore, if $\alpha \in (\underline{\alpha}, 1]$ *then*

$$\begin{aligned} q(\alpha) \quad = \quad & \frac{\delta(a-d)+\delta(1-\delta)(d-c-y)}{(1-\delta)\left[(b-a)-(d-c)\right]+\delta a-\delta(1-\delta)(x+y)} \\ & + \left(\frac{1}{\alpha}\right)\left[\frac{\delta a-\delta(1-\delta)x-\delta(a-d)-\left(1-\delta^2\right)(d-c)}{(1-\delta)\left[(b-a)-(d-c)\right]+\delta a-\delta(1-\delta)(x+y)}\right] \end{aligned} \tag{14}$$

In the preceding section, we established that $\alpha^t q\left(\alpha^t\right)$ converges to zero and thus is eventually decreasing over time. For affine PMPBE, we can now say that $\alpha^t q\left(\alpha^t\right)$ is monotonically declining over time, in which case the probability a player chooses C decreases with the length of time for which cooperative play has not yet occurred.

Theorem 5 *If* $q(\cdot)$ *is defined by (14) then* $\alpha q(\alpha)$ *is increasing in* α *and* $V(\alpha)$ *is increasing in* α.

While $\alpha q(\alpha)$ is increasing in α, $q(\alpha)$ need not be increasing in α everywhere, though we know that eventually it must be increasing in α since it converges to zero (when $\underline{\alpha} > 0$). We next show that when $d > c$ then $q(\alpha)$ is decreasing over time as lower probability is attached to players being type L (given only D has been chosen thus far). However, when $d = c$ then $q(\alpha)$ is, interestingly, independent of a player's beliefs as to the other player's type and thus is constant over time. Though it is still the case that α^t is declining, a type L player maintains the same probability of acting cooperatively.

Theorem 6 *If $q(\cdot)$ is defined by (14) then, for $\alpha > \underline{\alpha}$: i) if $d > c$ then $q(\alpha)$ is increasing in α; and ii) if $d = c$ then $q(\alpha) = q'$ for some $q' \in (0, 1)$.*

When $d = c$ - so a player is not harmed when choosing the cooperative action - the probability that a type L player chooses C is fixed at some positive value. Thus, if both players are type L then, almost surely, players will eventually achieve the collusive outcome. However, whether cooperative play ultimately emerges when $d > c$ is not so clear, as the probability of cooperation being initiated converges to zero. To examine this issue, define Q^T as the probability that players are still not colluding by the end of period T, conditional on both players being type L. Q^T is defined by

$$Q^T = \prod_{t=1}^{T} \left(1 - q^t\right)^2$$

where, given α^1, q^t is defined recursively by:

$$q^t = q\left(\alpha^t\right), t \geq 1;\ \alpha^t = \frac{\alpha^{t-1}\left(1 - q^{t-1}\right)}{1 - \alpha^{t-1} q^{t-1}}, t \geq 2.$$

The next result shows that, even when both players are type L, there is a positive probability that collusion never emerges even though they never give up trying (that is, they always choose C with positive probability).[19]

Theorem 7 *If $q(\alpha)$ is defined by (14) and $d > c$ then $\lim_{T\to\infty} Q^T > 0$.*

[19] Theorem 7 is true as long as $q(\alpha) = A + B\left(\frac{1}{\alpha}\right)$ for some A and B where $B < 0$ and $A + B < 1$.

If both players are type L then, in any period, there is always a positive probability that one of them will choose the cooperative action and thereby result in the emergence of collusion. It must then be true that a long sequence of choosing D is not a sufficiently pessimistic signal (that the other player is type L) which can only be the case if, as $\alpha^t \to \underline{\alpha}$, the probability that a type L player chooses C converges sufficiently fast to zero. But, as shown in the previous result, this also has the implication that the probability that two type L players start colluding in period t is going to zero sufficiently fast, which means collusion is not assured. If both players are willing and able to cooperate, there is a positive probability that they never do so though they never give up trying.

3.6 Examples

3.6.1 Example 1: Bertrand Price Game

Assume $b = 2a, d = c = 0$, and normalize so $a = 1$. This case approximates the Bertrand price game in which, for example, market demand is perfectly inelastic at two units with a maximum willingness to pay of 1, and firms have zero marginal cost. A firm's equilibrium strategy during the learning phase is[20] $q(\alpha) = \left(\sqrt{4\delta + 1} - 1\right) / \left(\sqrt{4\delta + 1} + 1\right)$. As one would expect, the probability of choosing C is higher when players are more patient.

3.6.2 Example 2: Bertrand Price Game with Relative Compensation

Let us modify the Bertrand price game so that managers - not owners - are repeatedly making price decisions and managerial compensation is based on relative performance. Specifically, a manager receives compensation equal to half of firm profit but, in the event that the other firm has higher profit, incurs a penalty equal to one-quarter of the rival firm's profit. The single-period payoff to a manager is then:

$$\text{Payoff of manager } i \text{ in period } t = \begin{cases} (1/2)\,\pi_i^t & \text{if } \pi_i^t \geq \pi_j^t \\ (1/2)\,\pi_i^t - (1/4)\,\pi_j^t & \text{if } \pi_i^t < \pi_j^t \end{cases}$$

[20]Derivations for all examples are available on request.

where π_i^t is the period t profit of firm i. If market demand is perfectly inelastic at two units with a maximum willingness to pay of 2 (and zero marginal cost) then the managers' payoff matrix is represented by: $a = 1, b = 2, c = -1, d = 0$. Equilibrium has:

$$q(\alpha) = \begin{cases} 0 & \text{if } \alpha \in \left[0, \frac{1-\delta^2}{2\delta-\delta^2}\right] \\ \frac{\left[\alpha\delta(2-\delta)-\left(1-\delta^2\right)\right]\left(\sqrt{2\delta}-1\right)}{\alpha\sqrt{2\delta}(2\delta-1)} & \text{if } \alpha \in \left(\frac{1-\delta^2}{2\delta-\delta^2}, 1\right] \end{cases}$$

If $\delta = .8$ then $\underline{\alpha} = .375$ and, for $\alpha > .375$, $q(\alpha) \simeq .335 - \frac{.126}{\alpha}$. If players have thus far always played D then, in each player updating their beliefs as to the other player's type, α^t will fall over time which then induces type L players to choose C with a lower probability. Assuming each firm initially assigns a 50% chance to its rival being type L, there is a 36% chance that collusion is never achieved.

3.6.3 Example 3: Asymmetric Bertrand Price Game

Consider the following generalization of Example 1 where the collusive outcome is now allowed to be asymmetric and $\gamma \in [1/2, 1)$.[21]

Asymmetric Bertrand Price Game

Player 2

Player 1		*Cooperate*	*Defect*
	Cooperate	$\gamma, 1-\gamma$	$0, 1$
	Defect	$1, 0$	$0, 0$

The collusive outcome gives player 1 a market share of γ which is at least 1/2. There is an affine PMPBE with

$$q_1 = \frac{\sqrt{\gamma(\gamma + 4\delta(1-\gamma))} - \gamma}{\sqrt{\gamma(\gamma + 4\delta(1-\gamma))} + \gamma}, \; q_2 = \frac{\sqrt{(1-\gamma)(1-\gamma+4\delta\gamma)} - (1-\gamma)}{\sqrt{(1-\gamma)(1-\gamma+4\delta\gamma)} + (1-\gamma)}$$

One can prove that q_1 is decreasing in γ and increasing in δ, and q_2 is increasing in γ and δ.

[21] A preliminary analysis suggests that many of the results in Sections 3 and 4 can be extended to when the Prisoners' Dilemma is asymmetric.

It might be expected that the player with the higher share of collusive profit would play C with a higher probability. However, when the share of collusive profit for player 1 (γ) is larger, the probability of playing C is actually higher for player 2 and lower for player 1. Since player 1 gains more by achieving cooperative play when γ is bigger, player 2 must be more likely to play C if player 1 is to be indifferent between playing C and D; and recall that D is more attractive when the other player is more likely to initiate cooperation. The player who benefits more from colluding is then less likely to take the first move in cooperating.

To explore the effect of asymmetry on the likelihood of collusion, consider the probability that collusion is initiated in any period:

$$1-(1-q_1)(1-q_2)=1-\frac{4}{\left(\sqrt{\frac{\gamma+4\delta(1-\gamma)}{\gamma}}+1\right)\left(\sqrt{\frac{1-\gamma+4\delta\gamma}{1-\gamma}}+1\right)}.$$

It is straightforward to show that it is increasing in γ, so collusion is more likely when the collusive outcome is more skewed to favor one firm.

As the equilibrium condition for the grim trigger strategy is $\delta \geq \gamma$, increasing asymmetry by raising γ makes collusion more difficult in the sense that the minimum discount factor is higher. However, conditional on the collusive outcome being sustainable, asymmetry reduces the expected time until collusion is achieved. In fact, as asymmetry becomes extreme, collusion is achieved immediately. Since $\lim_{\gamma\to 1} q_1(\alpha_1)=0$ and $\lim_{\gamma\to 1} q_2(\alpha_2)=1$ then $\lim_{\gamma\to 1} 1-(1-q_1)(1-q_2)=1$.[22]

3.7 Concluding Remarks

In practice, communication - either express or implicit - is essential to collusion. This we know from both experimental work and the many documented episodes of cartels. Communication can manifest itself in two ways: exchange of information and exchange of intentions. There is a limited amount of work in oligopoly theory on collusion and the exchange of information. In Athey and Bagwell (2001, 2008), firms have private information about their cost and exchange (costless) messages about cost, while in Hanazono and Yang (2007) and Gerlach (2009), firms have private signals on demand and seek to share that information.

[22]Keep in mind that as we let $\gamma \to 1$, we must have $\delta \to 1$ so that $\delta \geq \gamma$ is satisfied.

Then there is work in which sales or some other endogenous variable is private information and firms exchange messages for monitoring purposes; see Aoyagi (2002), Chan and Zhang (2009), and Harrington and Skrzypacz (2011).[23] Communication may also be used to resolve strategic uncertainty; specifically, in order to coordinate a move from a non-collusive to a collusive equilibrium. Here, intentions rather than hard information is being communicated.

Within the context of the equilibrium paradigm, the current paper sought to make progress on the tacit signaling of the intention to collude. In a sense, signaling in our model is part information (regarding a player's type) and part intentions (regarding cooperative play). Let us summarize our main findings. If the initial probability that players are capable of colluding is sufficiently high then, in any period, there is always the prospect of collusion emerging; no matter how long is there a history of failed collusion, beliefs as to players being cooperative types remain sufficiently high that it is worthwhile for them to continue to try to cooperate. This does not imply, however, that collusion is assured. For a wide class of situations, there is a positive probability that collusion never emerges. Players never give up trying to collude but they may also never succeed.

In terms of future work, one research direction is to allow a player's type to change over time, rather than remain fixed forever.[24] When a cooperative type raises price and does not receive a favorable response, it'll infer that its rival is an uncooperative type. In that case, it might be inclined to try again later on the hope that the rival's type has changed. But it may also be the case that a player who has previously failed to respond in kind to an invitation to collude will see itself as having the onus to initiate cooperation (in the event that its type changes) because its rival believes it is an uncooperative type. Now suppose players are currently engaged in cooperative play. A deviation by a player is part of equilibrium play and signals a change in a player's type to being uncooperative. Assuming persistence in types, the punishment of the deviator would have a certain credibility (beyond simply being an equilibrium) in that the other player believes there is little point in trying to cooperate. Indeed, non-cooperation may be the unique equilibrium. All this could put the burden on

[23] There is also an extensive game theory literature on the issue of private monitoring. See Compte (1998), Kandori and Matsushima (1998), Kandori (2002), Zheng (2008), and Obara (2009)

[24] Recent work by Escobar and Toikka (2009) provides a foundation for such an analysis.

the deviator to re-initiate cooperation. Even this cursory analysis suggests that a rich set of behavior could arise from allowing types to evolve stochastically over time.

References

[1] Aoyagi, Masaki, "Collusion in Dynamic Bertrand Oligopoly with Correlated Private Signals and Communication," *Journal of Economic Theory*, 102 (2002), 229-248.

[2] Atakan, Alp E. and Mehmet Ekmekci, "Reputation with Long Run Players," J. L. Kellogg Graduate School of Management, Northwestern University, working paper, October 2008.

[3] Atakan, Alp E. and Mehmet Ekmekci, "A Two-Sided Reputation Result with Long Run Players," J. L. Kellogg Graduate School of Management, Northwestern University, working paper, August 2009.

[4] Athey, Susan and Kyle Bagwell, "Optimal Collusion with Private Information," *RAND Journal of Economics*, 32 (2001), 428-465.

[5] Athey, Susan and Kyle Bagwell, "Collusion with Persistent Cost Shocks," *Econometrica*, 76 (2008), 493-540.

[6] Aumann, Robert J. and Sylvain Sorin, "Cooperation and Bounded Recall, *Games and Economic Behavior*, 1 (1989), 5-39.

[7] Bhaskar, V., "Informational Constraints and the Overlapping Generations Model: Folk and Anti-Folk Theorems," *Review of Economic Studies*, 65 (1998), 135-149.

[8] Bhaskar, V., "The Robustness of Repeated Game Equilibria to Incomplete Payoff Information," University of Essex, working paper, 2000.

[9] Bhaskar, V., George J. Mailath, and Stephen Morris, "Purification in the Infinitely-Repeated Prisoners' Dilemma," *Review of Economic Dynamics*, 11 (2008), 515-528.

[10] Borenstein, Severin, "Rapid Price Communication and Coordination: The Airline Tariff Publishing Case (1994)," in *The Antitrust Revolution*, 4th edition, John E. Kwoka, Jr. and Lawrence J. White, eds., New York: Oxford University Press, 2004.

[11] Chan, Jimmy and Wenzhang Zhang, "Price Wars, Side-payments, and Collusion Enforcement." Shanghai University of Finance and Economics, 2009.

[12] Chassang, Sylvain, "Building Routines: Learning, Cooperation, and Dynamics of Incomplete Relational Contracts," *American Economic Review*, 100 (2010), 448-465.

[13] Clark, Robert and Jean-François Houde, "Collusion with Asymmetric Retailers: Evidence from a Gasoline Price-Fixing Case," HEC Montréal, March 2011.

[14] Compte, Olivier, "Communication in Repeated Games with Imperfect Private Monitoring," *Econometrica*, 66 (1998), 597-626.

[15] Cripps, Martin W. and Jonathan P. Thomas, "Reputation and Perfection in Repeated Common Interest Games," *Games and Economic Behavior*, 18 (1997), 141-158.

[16] Cripps, Martin W., Eddie Dekel, and Wolfgang Pesendorfer, "Reputation with Equal Discounting in Repeated Games with Strictly Conflicting Interests," *Journal of Economic Theory*, 121 (2005), 259-272.

[17] Crawford, Vincent P. and Hans Haller, "Learning How to Cooperate: Optimal Play in Repeated Coordination Games," *Econometrica*, 58 (1990), 571-595.

[18] Dal Bó, Pedro, "Cooperation under the Shadow of the Future: Experimental Evidence from Infinitely Repeated Games," *American Economic Review*, 95 (2005), 1591-1604.

[19] Dixit, Avinash and Carl Shapiro, "Entry Dynamics with Mixed Strategies," in *The Economics of Strategic Planning*, L. G. Thomas, ed., Lexington, Mass.: Lexington Books, 1985.

[20] Escobar, Juan F. and Juuso Toikka, "A Folk Theorem with Markovian Private Information," Stanford University, November 2009.

[21] Farrell, Joseph, "Cheap Talk, Coordination, and Entry," *Rand Journal of Economics*, 18 (1987), 34-39.

[22] Gerlach, Heiko, "Stochastic Market Sharing, Partial Communication and Collusion," *International Journal of Industrial Organization*, 27 (2009), 655-666.

[23] Halac, Marina, "Relational Contracts and the Value of Relationships," *American Economic Review*, 102 (2012), 750-779.

[24] Hanazono, Makato and Huanxing Yang, "Collusion, Fluctuating Demand, and Price Rigidity," *International Economic Review*, 48 (2007), 483-515.

[25] Harrington, Joseph E. Jr. and Andrzej Skrzypacz, "Private Monitoring and Communication in Cartels: Explaining Recent Collusive Practices," *American Economic Review*, 101 (2011), 2425-2449.

[26] Kandori, Michihiro, "Introduction to Repeated Games with Private Monitoring," *Journal of Economic Theory*, 102 (2002), 1-15.

[27] Kandori, Michihiro and Hitoshi Matsushima, "Private Observation, Communication and Collusion," *Econometrica*, 66 (1998), 627-652.

[28] Kreps, David M., Paul Milgrom, John Roberts, and Robert Wilson, "Rational Cooperation in the Finitely Repeated Prisoners' Dilemma," *Journal of Economic Theory*, 27 (1982), 245-252.

[29] Mailath, George J. and Larry Samuelson, "Your Reputation Is Who You're Not, Not Who You'd Like To Be," CARESS Working Paper 98-11, August 1998.

[30] Mailath, George J. and Larry Samuelson, *Repeated Games and Reputations: Long-Run Relationships*, Oxford: Oxford University Press, 2006.

[31] Obara, Ichiro, "Folk Theorem with Communication," *Journal of Economic Theory*, 144 (2009), 120-134.

[32] Zheng, Bingyong, "Approximate Efficiency in Repeated Games with Correlated Private Signals," *Games and Economic Behavior*, 63 (2008), 406-416.

3.8 Appendix: Proofs

Proof of Theorem 1. To start, let us show that $\underline{\alpha} \in [0,1)$. $\underline{\alpha} \geq 0$ follows from $d \geq c$ and $a > d$. $\underline{\alpha} < 1$ if and only if

$$\delta\left[(1-\delta)(a-c)+\delta(a-d)\right] > \left(1-\delta^2\right)(d-c) \Leftrightarrow \delta > \frac{d-c}{a-c}.$$

Given (1) holds, a sufficient condition for $\delta > \frac{d-c}{a-c}$ is

$$\frac{b-a}{b-d} \geq \frac{d-c}{a-c} \Rightarrow (b-a)(a-c) \geq (d-c)(b-d) \Rightarrow b+c \geq a+d$$

which is true by assumption.

We need (9) to hold when firms are randomizing, which means when $\alpha > \underline{\alpha}$. That is the case if $\underline{\alpha} \geq \alpha^*$ which, after some manipulation, is equivalent to:

$$\frac{1+\delta}{\delta\left[(1-\delta)(a-c)+\delta(a-d)\right]} \geq \frac{1}{\delta(a-d)-(1-\delta)\left[(b-a)-(d-c)\right]}. \tag{15}$$

(15) holds as $\delta \to 1$. Thus, as long as δ is sufficiently close to one (which is a condition of Theorem 1) then as soon as one player chooses C, both players will optimally adopt the grim trigger strategy when they are type L.

Now let us move on to establishing stated properties on $q(\cdot)$. A player strictly prefers D to C if and only if:

$$\begin{aligned} & \alpha q\left(b+\frac{\delta a}{1-\delta}\right)+(1-\alpha q)\left(d+\delta V\left(\frac{\alpha(1-q)}{1-\alpha q}\right)\right) \\ > \;& \alpha q\left(\frac{a}{1-\delta}\right)+\alpha(1-q)\left(c+\frac{\delta a}{1-\delta}\right)+(1-\alpha)\left(c+\delta c+\frac{\delta^2 d}{1-\delta}\right). \end{aligned} \tag{16}$$

Note that $V(\alpha)$ has a lower bound of $\frac{d}{1-\delta}$ - as a player can assure itself of a payoff of at least $\frac{d}{1-\delta}$ by always choosing D - which then implies $V\left(\frac{\alpha(1-q)}{1-\alpha q}\right) \geq \frac{d}{1-\delta}$. Substituting $\frac{d}{1-\delta}$

for $V\left(\frac{\alpha(1-q)}{1-\alpha q}\right)$ and re-arranging, a sufficient condition for (16) is

$$\begin{aligned} & \alpha q\left(b+\frac{\delta a}{1-\delta}\right)+(1-\alpha q)\left(\frac{d}{1-\delta}\right)-\alpha q\left(\frac{a}{1-\delta}\right) \\ & -\alpha(1-q)\left(c+\frac{\delta a}{1-\delta}\right)-(1-\alpha)\left(c+\delta c+\frac{\delta^2 d}{1-\delta}\right) \\ > \; & 0. \end{aligned} \tag{17}$$

Take the derivative of the LHS of (17) with respect to q:

$$\begin{aligned} & \alpha\left(b+\frac{\delta a}{1-\delta}\right)-\alpha\left(\frac{d}{1-\delta}\right)-\alpha\left(\frac{a}{1-\delta}\right)+\alpha\left(c+\frac{\delta a}{1-\delta}\right) \\ = \; & \alpha\left[(b-a)-(d-c)\right]+\alpha\delta\left(\frac{a-d}{1-\delta}\right)>0, \end{aligned} \tag{18}$$

since $b-a \geq d-c$ and $a-d>0$. Hence, the difference between the payoff to D and the payoff to C is minimized when $q=0$. Thus, D is surely strictly preferred to C if (17) holds when $q=0$:

$$\frac{d}{1-\delta}>\alpha\left(c+\frac{\delta a}{1-\delta}\right)+(1-\alpha)\left(c+\delta c+\frac{\delta^2 d}{1-\delta}\right) \Rightarrow \tag{19}$$

$$\alpha<\frac{(1+\delta)(d-c)}{\delta\left(a-c+\frac{\delta(a-d)}{1-\delta}\right)}=\frac{\left(1-\delta^2\right)(d-c)}{\delta\left[(1-\delta)(a-c)+\delta(a-d)\right]}\;(\equiv \underline{\alpha})$$

Thus, if $\alpha<\underline{\alpha}$ then, in equilibrium, $q(\alpha)=0$.

To prove that $q(\underline{\alpha})=0$, suppose not. It follows from $q(\underline{\alpha})>0$ that $\frac{\underline{\alpha}(1-q(\underline{\alpha}))}{1-\underline{\alpha}q(\underline{\alpha})}<\underline{\alpha}$. The preceding analysis showed $q(\alpha)=0\ \forall\alpha<\underline{\alpha}$ and, since $q=0$ implies $\frac{\alpha(1-q)}{1-\alpha q}=\alpha$, then by stationary $q^t=0\ \forall\ t\geq t'$ when $\alpha^{t'}<\underline{\alpha}$. Hence,

$$V\left(\frac{\underline{\alpha}\left(1-q\left(\underline{\alpha}\right)\right)}{1-\underline{\alpha}q\left(\underline{\alpha}\right)}\right)=\frac{d}{1-\delta}. \tag{20}$$

For $q(\underline{\alpha})>0$, the expected payoff from choosing C must be at least as great as that from choosing D:

$$\begin{aligned} & \alpha q\left(\frac{a}{1-\delta}\right)+\alpha(1-q)\left(c+\frac{\delta a}{1-\delta}\right)+(1-\alpha)\left(c+\delta c+\frac{\delta^2 d}{1-\delta}\right) \\ \geq \; & \alpha q\left(b+\frac{\delta a}{1-\delta}\right)+(1-\alpha q)\left(\frac{d}{1-\delta}\right), \end{aligned} \tag{21}$$

where we used (20). However, note that the expressions in (21) are the same as those in (17). By our previous analysis, if $\alpha = \underline{\alpha}$ then (17) holds with equality when $q = 0$ and with strict inequality when $q > 0$. We conclude that (21) and $q(\underline{\alpha}) > 0$ are inconsistent and, therefore, $q(\underline{\alpha}) = 0$.

Next let us show: if $\alpha > \underline{\alpha}$ and $q(\alpha)$ is part of a PMPBE then $q(\alpha) \in (0,1)$. To prove $q(\alpha) > 0$, suppose not so $\exists \alpha' > \underline{\alpha}$ such that $q(\alpha') = 0$. By the preceding logic, $V(\alpha') = \frac{d}{1-\delta}$. In that case, the payoff to D is at least as great as that from C if and only if (19) holds with a weak inequality which the previous analysis showed that is the case if and only if $\alpha \leq \underline{\alpha}$. Therefore, if $\alpha > \underline{\alpha}$ then $q(\alpha) > 0$. To show that $q(\alpha) < 1$, suppose $q(\alpha) = 1$. The payoffs from C and D are:

$$\begin{aligned} \text{Play C} &: \quad \alpha\left(\frac{a}{1-\delta}\right) + (1-\alpha)\left(c + \delta c + \frac{\delta^2 d}{1-\delta}\right) \\ \text{Play D} &: \quad \alpha\left(b + \frac{\delta a}{1-\delta}\right) + (1-\alpha)\left(d + \frac{\delta d}{1-\delta}\right), \end{aligned}$$

Since choosing D yields a strictly higher payoff $\forall \alpha \in (0,1]$, it follows that $q(\alpha)$ must be bounded below 1 $\forall \alpha \in [0,1]$. Therefore, $q(\alpha) < 1$ $\forall \alpha \in (0,1]$ and, furthermore, $\lim_{\alpha \to 1} q(\alpha) < 1$. Finally, Theorem 4 proves by construction that there exists $q(\cdot) : (\underline{\alpha}, 1] \to (0,1)$ for which the expected payoffs from C and D are equalized and thus is part of a PMPBE.

To complete the proof, let us show the properties on $V(\cdot)$ are true, given the properties on $q(\cdot)$ hold. First note that, in equilibrium, $V : [0,1] \to \left[\frac{d}{1-\delta}, \frac{a}{1-\delta}\right]$, as $V(\alpha)$ has a lower bound of $\frac{d}{1-\delta}$ and $\frac{a}{1-\delta}$ is an upper bound because the highest average symmetric payoff is a. If $q(\alpha) = 0$ then type L players play D for sure in the current period and since $\frac{\alpha(1-q(\alpha))}{1-\alpha q} = \alpha$ then the same is true for all ensuing periods; hence, by stationarity, if $q(\alpha) = 0$ then $V(\alpha) = \frac{d}{1-\delta}$. To show that $V(\alpha) \in \left(\frac{d}{1-\delta}, \frac{a}{1-\delta}\right)$ when $\alpha \in (\underline{\alpha}, 1)$, note that $q(\alpha) \in (0,1)$ implies $V(\alpha) = W^C(\alpha) = W^D(\alpha)$. $\frac{d}{1-\delta}$ is a lower bound on $V(\alpha)$ for all α since at least that value can be achieved by choosing D in every period. Using the payoff from choosing D, we have:

$$\begin{aligned} V(\alpha) &= \alpha q \left(b + \frac{\delta a}{1-\delta}\right) + (1-\alpha q)\left(d + \delta V\left(\frac{\alpha(1-q)}{1-\alpha q}\right)\right) \\ &\geq \alpha q \left(b + \frac{\delta a}{1-\delta}\right) + (1-\alpha q)\left(d + \delta \frac{d}{1-\delta}\right) \\ &> \frac{d}{1-\delta} + \alpha q \left(\frac{a-d}{1-\delta}\right) > \frac{d}{1-\delta} \end{aligned}$$

since $b > a > d$. Using the payoff from choosing C, we have:

$$\begin{aligned} V(\alpha) &= \alpha q \left(\frac{a}{1-\delta}\right) + \alpha(1-q)\left(c + \frac{\delta a}{1-\delta}\right) + (1-\alpha)\left(c + \delta c + \frac{\delta^2 d}{1-\delta}\right) \\ &= \frac{a}{1-\delta} - \alpha(1-q)(a-c) - (1-\alpha)\left(\frac{a}{1-\delta} - c - \delta c - \frac{\delta^2 d}{1-\delta}\right) < \frac{a}{1-\delta}, \end{aligned}$$

since $a > c, d$. This establishes the properties on $V(\cdot)$. ■

Proof of Theorem 2. As a preliminary result, let us first show: if $\alpha^1 > \underline{\alpha}$ then $\underline{\alpha}$ is a lower bound of the sequence $\{\alpha^t\}$. To do so, we'll show: if $\alpha > \underline{\alpha}$ then $\frac{\alpha(1-q(\alpha))}{1-\alpha q(\alpha)} > \underline{\alpha}$; recall that $\alpha^{t+1} = \frac{\alpha^t(1-q(\alpha^t))}{1-\alpha^t q(\alpha^t)}$. Suppose not so that $\exists \alpha' > \underline{\alpha}$ such that $\frac{\alpha'(1-q(\alpha'))}{1-\alpha' q(\alpha')} \leq \underline{\alpha}$. By the proof of Theorem 1, $V\left(\frac{\alpha'(1-q(\alpha'))}{1-\alpha' q(\alpha')}\right) = \frac{d}{1-\delta}$ and, from (8), we have:

$$\alpha' q(\alpha') = \frac{\delta\left(\frac{a-d}{1-\delta}\right) - [1 + \delta(1-\alpha')](d-c) - (1-\alpha')\frac{\delta(a-d)}{1-\delta}}{\delta\left(\frac{a-d}{1-\delta}\right) + (b-a) - (d-c)} \tag{22}$$

We've made the supposition $\underline{\alpha} \geq \frac{\alpha'(1-q(\alpha'))}{1-\alpha' q^*(\alpha')}$ which is equivalent to

$$\alpha' q(\alpha') \geq \frac{\alpha' - \underline{\alpha}}{1-\underline{\alpha}}. \tag{23}$$

Substituting (22) into (23),

$$\begin{aligned} &\frac{\alpha'\delta(a-d) - (1-\delta)(d-c) - \delta(1-\delta)(1-\alpha)(d-c)}{\delta(a-d) + (1-\delta)[(b-a)-(d-c)]} \\ \geq\ &\frac{\alpha\delta(1-\delta)(a-c) + \alpha\delta^2(a-d) - \left(1-\delta^2\right)(d-c)}{\delta[(1-\delta)(a-c) + \delta(a-d)] - \left(1-\delta^2\right)(d-c)} \end{aligned}$$

As the numerators are equal and positive (since they equal $\alpha' - \underline{\alpha}$) then the inequality holds if and only if

$$\begin{aligned}\delta\left[(1-\delta)(a-c)+\delta(a-d)\right]-\left(1-\delta^2\right)(d-c) &\geq \delta(a-d)+(1-\delta)\left[(b-a)-(d-c)\right] \Rightarrow \\ 0 &\geq (1-\delta)(b-a)\end{aligned}$$

which is not true. Hence, $\nexists \alpha' > \underline{\alpha}$ such that $\frac{\alpha'(1-q(\alpha'))}{1-\alpha' q(\alpha')} \leq \underline{\alpha}$; therefore, if $\alpha' > \underline{\alpha}$ then $\frac{\alpha'(1-q(\alpha'))}{1-\alpha' q(\alpha')} > \underline{\alpha}$.

Given that $\alpha^1 > \underline{\alpha}$ implies $\alpha^t > \underline{\alpha}$ $\forall t$, it follows from Theorem 1 that $q\left(\alpha^t\right) > 0$ $\forall t$ which further implies $\left\{\alpha^t\right\}$ is strictly decreasing. To complete the proof of part (i) of this theorem, it needs to be shown: if $\alpha^1 > \underline{\alpha}$ then $\lim_{t\to\infty} \alpha^t = \underline{\alpha}$. By Bayes rule,

$$\alpha^{t+1} = \alpha^t \left(\frac{1-q^t}{1-\alpha^t q^t}\right) \Rightarrow \alpha^{t+1} \leq \alpha^t.$$

Since it has already been shown that $\underline{\alpha}$ is a lower bound of the sequence $\{\alpha^t\}$, $\{\alpha^t\}$ has a limit and it is sufficient to show that $\underline{\alpha}$ is the infimum of $\{\alpha^t\}$. Suppose not, and let $\alpha' > \underline{\alpha}$ be the infimum of $\{\alpha^t\}$. Hence, as $\alpha^t \to \alpha'$ then $\alpha^{t+1} \to \alpha^t$, which means $q^t \to 0$, and $q^t \to 0$ implies $V(\alpha^{t+1}) \to \frac{d}{1-\delta}$. But we know from the proof of Theorem 1 that the payoff to D is the same as the payoff from C if and only if $\alpha^t \to \underline{\alpha}$, which contradicts $\alpha^t \to \alpha'$ and $\alpha' > \underline{\alpha}$. Therefore, $\lim_{t\to\infty} \alpha^t = \underline{\alpha}$, for $\alpha^1 > \underline{\alpha}$.

Next let us show part (ii): $\lim_{\alpha\downarrow\underline{\alpha}} q(\alpha) = 0$ when $\underline{\alpha} > 0$. As it has already been proven that $\alpha^1 > \underline{\alpha}$ implies $\lim_{t\to\infty} \alpha^t = \underline{\alpha}$, it follows that

$$\lim_{\alpha\downarrow\underline{\alpha}} \frac{\alpha(1-q(\alpha))}{1-\alpha q(\alpha)} = \underline{\alpha}\,(>0),$$

which implies $\lim_{\alpha\downarrow\underline{\alpha}} q(\alpha) = 0$.

Finally, it is easy to prove part (iii): $\lim_{t\to\infty} \alpha^t q\left(\alpha^t\right) = 0$. If $\alpha^1 \leq \underline{\alpha}$ then $q\left(\alpha^t\right) = 0$ $\forall t$ and therefore $\lim_{t\to\infty} \alpha^t q\left(\alpha^t\right) = 0$. If $\alpha^1 > \underline{\alpha} > 0$ then, by the other results of Theorem 2, $\lim_{t\to\infty} \alpha^t = \underline{\alpha}$ and $\lim_{\alpha\downarrow\underline{\alpha}} q(\alpha) = 0$ which implies $\lim_{t\to\infty} \alpha^t q\left(\alpha^t\right) = 0$. If $\alpha^1 > \underline{\alpha} = 0$ then $\lim_{t\to\infty} \alpha^t = 0$ which implies $\lim_{t\to\infty} \alpha^t q\left(\alpha^t\right) = 0$. ■

Proof of Theorem 4. First note that, by the same argument in the proof of Theorem 1, if δ is sufficiently high then $\alpha \geq \alpha^*$.

Re-arranging (8), an equilibrium $q(\cdot)$ is defined by

$$\begin{aligned} & \alpha q\left[(b-a)-(d-c)\right]+(1-\alpha)\frac{\delta(a-d)}{1-\delta}+(d-c)+\delta(1-\alpha)(d-c) \\ = \; & \delta(1-\alpha q)\left[\frac{a}{1-\delta}-V\left(\frac{\alpha(1-q)}{1-\alpha q}\right)\right] \end{aligned} \tag{24}$$

Conjecturing that the value function is linear in α,

$$V(\alpha)=x+y\alpha, \tag{25}$$

substitute (25) into (24).

$$\begin{aligned} & \alpha q\left[(b-a)-(d-c)\right]+(1-\alpha)\frac{\delta(a-d)}{1-\delta}+(d-c)+\delta(1-\alpha)(d-c) \\ = \; & \delta(1-\alpha q)\left[\frac{a}{1-\delta}-x-y\left(\frac{\alpha(1-q)}{1-\alpha q}\right)\right] \Rightarrow \end{aligned} \tag{26}$$

$$\begin{aligned} \alpha q \;=\; & \alpha\left[\frac{\delta(a-d)+\delta(1-\delta)(d-c-y)}{(1-\delta)\left[(b-a)-(d-c)\right]+\delta a-\delta(1-\delta)(x+y)}\right] \\ & +\frac{\delta a-\delta(1-\delta)x-\delta(a-d)-\left(1-\delta^2\right)(d-c)}{(1-\delta)\left[(b-a)-(d-c)\right]+\delta a-\delta(1-\delta)(x+y)} \end{aligned} \tag{27}$$

Thus, αq is affine in α if the value function is affine in α. As a player is indifferent between playing C and D, the value can be given by the payoff to choosing C for sure:

$$V(\alpha)=\alpha q(a-c)+\frac{\alpha\delta(a-d)}{1-\delta}+c+\frac{\delta d}{1-\delta}-\delta(1-\alpha)(d-c).$$

The value function is affine in αq and, since αq is affine in α, $V(\alpha)$ is affine in α.

The next step is to show that there exist unique values for x and y. Using the payoff to playing C, in equilibrium the value function equals:

$$V(\alpha)=\alpha q(a-c)+c+\frac{\alpha\delta(a-d)}{1-\delta}+\frac{\delta d}{1-\delta}-\delta(1-\alpha)(d-c)$$

$$\begin{aligned}
&= \alpha\left[\frac{\delta\left(a-c\right)\left[\left(a-d\right)+\left(1-\delta\right)\left(d-c-y\right)\right]}{\left(1-\delta\right)\left[\left(b-a\right)-\left(d-c\right)\right]+\delta a-\delta\left(1-\delta\right)\left(x+y\right)}+\right. \\
&\quad \left.\frac{\delta\left(a-d\right)}{1-\delta}+\delta\left(d-c\right)\right] \\
&\quad +\left(a-c\right)\left[\frac{\delta a-\delta\left(1-\delta\right)x-\delta\left(a-d\right)-\left(1-\delta^2\right)\left(d-c\right)}{\left(1-\delta\right)\left[\left(b-a\right)-\left(d-c\right)\right]+\delta a-\delta\left(1-\delta\right)\left(x+y\right)}\right] \\
&\quad +c+\frac{\delta d}{1-\delta}-\delta\left(d-c\right)
\end{aligned} \tag{28}$$

Equating coefficients between (25) and (28), we have

$$\begin{aligned}
x &= \left(a-c\right)\left[\frac{\delta a-\delta\left(1-\delta\right)x-\delta\left(a-d\right)-\left(1-\delta^2\right)\left(d-c\right)}{\left(1-\delta\right)\left[\left(b-a\right)-\left(d-c\right)\right]+\delta a-\delta\left(1-\delta\right)\left(x+y\right)}\right] \\
&\quad +c+\frac{\delta d}{1-\delta}-\delta\left(d-c\right)
\end{aligned} \tag{29}$$

$$y=\frac{\delta\left(a-c\right)\left[\left(a-d\right)+\left(1-\delta\right)\left(d-c-y\right)\right]}{\left(1-\delta\right)\left[\left(b-a\right)-\left(d-c\right)\right]+\delta a-\delta\left(1-\delta\right)\left(x+y\right)}+\frac{\delta\left(a-d\right)}{1-\delta}+\delta\left(d-c\right) \tag{30}$$

To show that there is a unique solution to (29)-(30), define $z\equiv x+y$ and note that:

$$z=x+y=V(1)=W^C(1)=Q(a-c)+\frac{\delta\left(a-d\right)}{1-\delta}+c+\frac{\delta d}{1-\delta},$$

where $Q=q(1)$. Simplifying the preceding equation gives:

$$z=Q(a-c)+\frac{\delta a}{1-\delta}+c. \tag{31}$$

If we can show that there exists a unique $Q\in(0,1)$ satisfying the equilibrium condition (26) when $\alpha=1$, then $z=x+y=V(1)$ is unique.

Evaluating (26) at $\alpha=1$, we have:

$$Q\left[\left(b-a\right)-\left(d-c\right)\right]+\left(d-c\right)=\delta\left(1-Q\right)\left[\frac{a}{1-\delta}-x-y\left(\frac{1-Q}{1-Q}\right)\right]\Rightarrow$$

$$\delta\left(a-c\right)Q^2-\left[2\delta\left(a-c\right)+\left(b-a\right)-\left(d-c\right)\right]Q+\left[\delta\left(a-c\right)-\left(d-c\right)\right]=0.$$

This quadratic has two solutions:

$$Q = \frac{2\delta (a-c) + (b-a) - (d-c) \pm \sqrt{\Omega}}{2\delta (a-c)},$$

where

$$\Omega \equiv [2\delta (a-c) + (b-a) - (d-c)]^2 - 4\delta (a-c) [\delta (a-c) - (d-c)] \tag{32}$$

$$\begin{aligned} &= [(b-a) - (d-c)]^2 + 4\delta (a-c)(b-a) \\ &> 0 \end{aligned}$$

since $b > a > c$. Hence, the two solutions are real. Next note that the bigger root exceeds one:

$$Q^b = 1 + \frac{(b-a) - (d-c) + \sqrt{\Omega}}{2\delta (a-c)} > 1.$$

Thus, we only need to show that the smaller root falls in $(0, 1)$.

$$Q^s = 1 + \frac{(b-a) - (d-c) - \sqrt{\Omega}}{2\delta (a-c)} < 1$$

if and only if

$$\begin{aligned} (b-a) - (d-c) &< \sqrt{\Omega} \Leftrightarrow [(b-a) - (d-c)]^2 < \Omega \Leftrightarrow \\ [(b-a) - (d-c)]^2 &< [(b-a) - (d-c)]^2 + 4\delta (a-c)(b-a) \\ &\Leftrightarrow 0 < 4\delta (a-c)(b-a), \end{aligned}$$

therefore, $Q^s < 1$. $Q^s > 0$ if and only if

$$[2\delta (a-c) + (b-a) - (d-c)]^2 > \Omega.$$

From (32), the preceding condition is equivalent to

$$4\delta (a-c) [\delta (a-c) - (d-c)] > 0,$$

which holds since $\delta > \frac{d-c}{a-c}$. The last property follows from $\delta > \frac{b-a}{b-d} \geq \frac{d-c}{a-c}$.

There then exists a unique $Q \in (0,1)$, and $z = x + y = V(1)$ is unique since it is linear in Q. In addition, plugging Q^s in (31) gives

$$\begin{aligned} z &= \frac{2\delta (a-c) + (b-a) - (d-c) - \sqrt{\Omega}}{2\delta} + \frac{\delta a}{1-\delta} + c \\ &= \frac{2a\delta + (1-\delta)\left[(b-a) - (d-c) - \sqrt{\Omega}\right]}{2\delta (1-\delta)}. \end{aligned}$$

To close the model, use the initial condition $V(\underline{\alpha}) = \frac{d}{1-\delta}$, which takes the form:

$$x = \frac{d}{1-\delta} - y \frac{\left(1-\delta^2\right)(d-c)}{\delta \left[(1-\delta)(a-c) + \delta (a-d)\right]}.$$

x^* is then the unique solution to

$$x^* = \frac{d}{1-\delta} - (z - x^*) \frac{\left(1-\delta^2\right)(d-c)}{\delta \left[(1-\delta)(a-c) + \delta (a-d)\right]},$$

and y^* is the unique solution to: $y^* = z - x^*$. This completes the proof that there is a unique affine PMPBE. Finally, solving for q from (27) gives us (14).

This construction of an equilibrium $q(\cdot)$ was the basis for the existence statement in Theorem 1. As that statement also had $q(\alpha) \in (0,1)$ if $\alpha \in (\underline{\alpha}, 1]$ then we need (14) to satisfy that condition as well. Given that q is increasing in α (see Theorem 6), it need only be shown that $q(\underline{\alpha}) = 0$ and $q(1) < 1$. Those two properties are straightforward to establish by substituting $\underline{\alpha}$ and 1, respectively, into (14). ■

Proof of Theorem 5. Since the equilibrium probability of choosing C is

$$\begin{aligned} \alpha q(\alpha) &= \alpha \left[\frac{\delta (a-d) + \delta (1-\delta)(d-c-y)}{(1-\delta)\left[(b-a) - (d-c)\right] + \delta a - \delta (1-\delta)(x+y)}\right] \\ &\quad + \left[\frac{\delta a - \delta (1-\delta) x - \delta (a-d) - \left(1-\delta^2\right)(d-c)}{(1-\delta)\left[(b-a) - (d-c)\right] + \delta a - \delta (1-\delta)(x+y)}\right] \end{aligned}$$

then $\alpha q(\alpha)$ is increasing in α if and only if

$$\frac{\delta(a-d)+\delta(1-\delta)(d-c-y)}{(1-\delta)[(b-a)-(d-c)]+\delta a-\delta(1-\delta)(x+y)}>0. \tag{33}$$

By assumption $(b-a)-(d-c)\geq 0$, and $V(1)<\frac{a}{1-\delta}$ implies

$$\frac{a}{1-\delta}-(x+y)>0. \tag{34}$$

Thus, (33) is true if and only if the numerator is positive:

$$\frac{(a-d)}{1-\delta}+(d-c)>y. \tag{35}$$

Suppose (35) was not true. From (30), we have

$$y=\frac{\delta(a-c)[(a-d)+(1-\delta)(d-c-y)]}{(1-\delta)[(b-a)-(d-c)]+\delta a-\delta(1-\delta)(x+y)}+\frac{\delta(a-d)}{1-\delta}+\delta(d-c). \tag{36}$$

If (35) is not true then the first term of (36) is non-positive, but then (36) implies

$$y\leq\delta\left[\frac{(a-d)}{1-\delta}+(d-c)\right]<\frac{(a-d)}{1-\delta}+(d-c)$$

which contradicts the supposition that (35) is not true. From this contradiction, we conclude (35) and thus $\alpha q(\alpha)$ is increasing in α.

To show that $V(\alpha)$ is increasing in α, recall that

$$V(\alpha)=\alpha q(\alpha)(a-c)+\frac{\alpha\delta(a-d)}{1-\delta}+c+\frac{\delta d}{1-\delta}-\delta(1-\alpha)(d-c).$$

That $\alpha q(\alpha)$ is increasing in α delivers the result. ■

Proof of Theorem 6. For $\alpha\leq\underline{\alpha}$, $q(\alpha)=0$, so it is non-decreasing in α for $\alpha\in[0,\underline{\alpha}]$.

From hereon, suppose $\alpha > \underline{\alpha}$ so that

$$
\begin{aligned}
q(\alpha) &= \frac{\delta(a-d)+\delta(1-\delta)(d-c-y)}{(1-\delta)[(b-a)-(d-c)]+\delta a-\delta(1-\delta)(x+y)} \\
&\quad + \left(\frac{1}{\alpha}\right)\left[\frac{\delta a-\delta(1-\delta)x-\delta(a-d)-\left(1-\delta^2\right)(d-c)}{(1-\delta)[(b-a)-(d-c)]+\delta a-\delta(1-\delta)(x+y)}\right]
\end{aligned}
$$

Thus, $q(\alpha)$ is increasing in α if and only if

$$
\left[\frac{\delta a-\delta(1-\delta)x-\delta(a-d)-\left(1-\delta^2\right)(d-c)}{(1-\delta)[(b-a)-(d-c)]+\delta a-\delta(1-\delta)(x+y)}\right]<0 \tag{37}
$$

The denominator of the LHS of (37) is positive because $b-a \geq d-c$ by assumption and $\frac{a}{1-\delta} > x+y$ as shown in (34). Thus, (37) is true if and only if the numerator is negative:

$$
\delta a-\delta(1-\delta)x-\delta(a-d)-\left(1-\delta^2\right)(d-c)<0. \tag{38}
$$

Suppose (38) was not true. From (29), we would then have $x \geq c+\frac{\delta d}{1-\delta}-\delta(d-c)$ which implies

$$
\begin{aligned}
&\delta a-\delta(1-\delta)x-\delta(a-d)-\left(1-\delta^2\right)(d-c) \\
\leq\ &\delta a-\delta(1-\delta)\left(c+\frac{\delta d}{1-\delta}-\delta(d-c)\right)-\delta(a-d)-\left(1-\delta^2\right)(d-c)
\end{aligned} \tag{39}
$$

By rearranging terms, the RHS of (39) is equivalent to

$$
-(1-\delta)(1-\delta^2)(d-c) \tag{40}
$$

which is negative if and only if $d > c$. Hence, the LHS of (38) is negative for $d > c$, which contradicts the supposition that (38) is not true. From this contradiction, we conclude (38) is true for $d > c$. Namely, $q(\alpha)$ is increasing in α for $d > c$.

If $d = c$, (40) implies

$$\delta a - \delta(1-\delta)x - \delta(a-d) - \left(1-\delta^2\right)(d-c) = 0 \Rightarrow \frac{\partial q(\alpha)}{\partial \alpha} = 0.$$

■

Proof of Theorem 7. First note that if $\alpha^1 \leq \underline{\alpha}$ then $q^t = 0\ \forall\ t$ in which case $Q^T = 1$. From hereon, assume $\alpha^1 \in (\underline{\alpha}, 1)$. If $d > c$ then, with the affine PMPBE, $q(\alpha) = A + B\left(\frac{1}{\alpha}\right)$ for some A and B where $B < 0$ and $A + B < 1$. Then

$$\alpha^t = \frac{\alpha^{t-1}\left(1-q^{t-1}\right)}{1-\alpha^{t-1}q^{t-1}} = \frac{\alpha^{t-1}\left(1-A-\frac{B}{\alpha^{t-1}}\right)}{1-\alpha^{t-1}\left(A+\frac{B}{\alpha^{t-1}}\right)}$$

$$q^t = A + B\left(\frac{1}{\alpha^t}\right) = A + B\left(\frac{1-\alpha^{t-1}\left(A+\frac{B}{\alpha^{t-1}}\right)}{\alpha^{t-1}\left(1-A-\frac{B}{\alpha^{t-1}}\right)}\right) \tag{41}$$

Since $B \neq 0$, we can invert $q^{t-1} = A + B\left(\frac{1}{\alpha^{t-1}}\right)$ to derive $\alpha^{t-1} = \frac{B}{q^{t-1}-A}$. Insert this expression in (41),

$$q^t = A + B\left(\frac{1-\left(\frac{B}{q^{t-1}-A}\right)\left(A+\frac{B}{\frac{B}{q^{t-1}-A}}\right)}{\left(\frac{B}{q^{t-1}-A}\right)\left(1-A-\frac{B}{\frac{B}{q^{t-1}-A}}\right)}\right) = q^{t-1}\left(\frac{1-A-B}{1-q^{t-1}}\right) \tag{42}$$

By $B < 0$ and $\alpha^t < 1$, we have $A + B > A + B\left(\frac{1}{\alpha^t}\right) = q^t$, $\forall\ t$. By $B < 0$ and that α^t decreasing over time, we have that q^t decreasing over time. Hence, $1 - q^1 \leq 1 - q^{t-1}$, $\forall$ $t > 2$. Therefore, $q^t \leq \left(\frac{1-A-B}{1-q^1}\right)q^{t-1}$. As this holds for all t, it implies

$$q^t \leq \left(\frac{1-A-B}{1-q^1}\right)^{t-1} q^1 = \nu^{t-1}q,$$

where $\nu \equiv \left(\frac{1-A-B}{1-q^1}\right) \in (0,1)$. Hence,

$$\prod_{t=1}^{T}\left(1-q^t\right)^2 > \left[\prod_{t=1}^{T}\left(1-\nu^{t-1}q\right)\right]^2.$$

To prove this theorem, it is then sufficient to show $\lim_{T\to\infty} \prod_{t=1}^{T} \left(1-\nu^{t-1}q\right) > 0$, or equivalently $\lim_{T\to\infty} \prod_{t=1}^{T} \left(1-\nu^{t}q\right) > 0$, which, because $q \in (0,1)$, is true if $\lim_{T\to\infty} \prod_{t=1}^{T} \left(1-\nu^{t}\right) > 0$, which is equivalent to $\sum_{t=1}^{\infty} \log\left(1-\nu^{t}\right) > -\infty$. Since $\nu \in (0,1)$ then

$$
\begin{aligned}
& \sum_{t=1}^{\infty} \log\left(1-\nu^{t}\right) \\
= & -\left[\left(\nu + \frac{\nu^2}{2} + \frac{\nu^3}{3} + ...\right) + \left(\nu^2 + \frac{\nu^4}{2} + \frac{\nu^6}{3} + ...\right) + ... + \left(\nu^t + \frac{\nu^{2t}}{2} + \frac{\nu^{3t}}{3} + ...\right) + ...\right] \\
= & -\left[\left(\nu + \nu^2 + \nu^3 + ...\right) + \frac{1}{2}\left(\nu^2 + \nu^4 + \nu^6 + ...\right) + \frac{1}{3}\left(\nu^3 + \nu^6 + \nu^9 + ...\right) + ...\right] \\
= & -\left(\frac{\nu}{1-\nu} + \frac{1}{2}\frac{\nu^2}{1-\nu^2} + \frac{1}{3}\frac{\nu^3}{1-\nu^3} + ...\right) = -\frac{\nu}{1-\nu}\left(1 + \frac{1}{2}\frac{\nu}{1+\nu} + \frac{1}{3}\frac{\nu^2}{1+\nu+\nu^2} + ...\right) \\
\geq & -\frac{\nu}{1-\nu}\left(1 + \nu + \nu^2 + ...\right) = -\frac{\nu}{\left(1-\nu\right)^2} > -\infty
\end{aligned}
$$

■

4 An Empirical Framework for Exclusive Discount

4.1 Introduction

Rather than directly constraining buyers' purchases, exclusive dealing sometimes takes a more subtle form: discounting with an exclusive requirement. For exclusive discount, buyers receive the discount price only if they make a certain proportion of their total purchase (market-share requirement) or a certain amount of purchases (volume requirement[1]) from the supplier. If buyers do not agree or do not meet the requirement, they have to pay a higher noncompliance price.[2] When the market-share requirement is 100%, the discount is perfectly exclusive. When in addition the noncompliance price is set to infinity, it gives the special case of perfect exclusion. This paper focuses on building an empirical framework for perfect exclusive discount. Exclusive discount is more complicated when applied to bundled products. The discount may only apply to buyers who obtain the bundle from the supplier (hereafter "bundled discount") or it may also apply to buyers who purchase a subset of products in the bundle as long as they restrict their purchase from rival suppliers (hereafter "exclusive discount" in a narrow sense). As shown in this paper, different forms of discounts may lead to opposite conclusions in consumer welfare analysis. For example, in some cases, exclusive discount may benefit consumers while bundled discount hurts them under the same market conditions. Thus, it is crucial to distinguish forms of discount in antitrust practice.

Exclusive discount has raised various antitrust concerns in recent years.[3] Despite its name, exclusive discount is not necessarily a true discount. The so-called "discount" is relative to the noncompliance or stand-alone prices rather than to the but-for prices. The major antitrust concern with exclusive discount is its potential foreclosure effect: a market-share requirement or volume target might hurt competitors by foreclosing a certain portion

[1] Though not explicitly so, a volume requirement is generally exclusive as it decreases residual demand for competitors' products.

[2] Noncompliance prices are not always higher than the bundled "discount." Rey and Tirole (2013) provides an example where bundled prices are higher then the sum of stand-alone prices.

[3] Important examples includes SmithKline v. Eli Lilly, Ortho v. Abbot, Concord Boat Corp. v. Brunswick Corp, Virgin Atlantic v. British Airways, and LePage's v. 3M. See Greenlee and Reitman (2005) and Kobayashi (2005a) for a review.

of market demand. Even if competitors are not foreclosed from the market, exclusive discount may serve as a tool to segregate the markets and reduce competition. Bundled discount also raises concerns in price discrimination against buyers who are only interested in some of the products in the bundle. In addition, as discussed in Elhauge (2009a, 2009b), when the magnitude (either in percentage or absolute value) of discount is fixed, exclusive discount might discourage the supplier from competing for free buyers.

Despite of these potential anticompetitive effects, an exclusive discount might also be procompetitive. For example, exclusivity might help in avoiding free-riding, reducing uncertainties, and generating cost efficiencies. More importantly, exclusive discount might reflect true price discount and thus enhance competition and benefit consumers even absent any of these procompetitive effects. In this case, exclusive discount allows the firm to charge different prices for different bundles. Thus, it can price more aggressively in one bundle without cutting prices for other bundles. Hence, whether a specific exclusive discount is detrimental to consumers or not should be examined case by case rather than be declared legal or illegal *per se*.

There is a gap between economic theory and antitrust practice in the sense that there exist few empirical works that are both rooted in current economic theory and directly applicable for case-by-case antitrust practice.[4] Economic theory generally assumes homogeneous product together with other simplifications. It is generally unclear how to apply these models to a specific case and how to estimate the demand and cost functions. On the other hand, antitrust practice often relies on rule-of-thumb tests that lack a complete economic theory foundation or do not necessarily target consumer welfare. For example, price–cost tests, such as the Ortho test, generally evaluate whether the discount prices are above average cost so that it will not foreclose an equal competitor. However, these tests are inappropriate because they focus only on the supply side of the market. The antitrust law is intended to protect competition rather than competitors and the ultimate standard should be based on impact on consumer welfare. Hence, a proper empirical analysis should provide predictions on consumer welfare.[5]

[4] Pereira et al. (2012) is one of the few papers using current economic theory to model demand and define market for bundled products.

[5] Some theoretical works also provide tests that target consumer welfare. For example, Greenlee et al.

This paper attempts to fill the gap between economic theory and antitrust practice for bundled and exclusive discount by developing empirical frameworks directly applicable to case studies. The frameworks are based on the discrete-choice models popular in industrial-organization literature since Berry (1994) and Berry et al. (1995) employed them. The paper shows that by defining product bundles as choice alternatives and differentiating between products and alternatives, the discrete-choice model can be applied to estimate demand with bundled or exclusive discount. The framework also clarifies the difference between a bundled discount and an exclusive discount. The numerical examples demonstrate that these differences are important as there are cases where bundle discount hurts consumers while exclusive discount makes them better off and vice versa. The framework admits a variety of variations. To name a few, it applies to cases where consumers are allowed to buy products in the bundle from different firms and to cases where they are not. It also applies when not all firms offer all products in the bundle. The framework assumes away potential procompetitive effects related to free-riding, uncertainty and cost efficiencies, which are usually specific to industry characteristics and difficult to evaluate in practice. Thus, if any of these procompetitive effects is believed to be important to an industry, applying the framework provides conservative conclusions on the benefit of exclusive discount.

With the estimated-demand system, counterfactual analysis can be performed to evaluate the impact on consumers and producers of an environmental or policy change, for example, disallowing exclusive discount or a new entrant. As the difference between products and alternatives complicates the computations in counterfactuals, numerical examples and solution steps (in the Appendix) are provided to demonstrate how to apply the framework. The numerical examples also show that welfare effects of bundled or exclusive discount are case-specific and can be obtained by applying the framework to a specific case.

In conclusion, the paper provides a general empirical framework for a variety of exclusive discounts, including bundled-product markets and commitment in discounts. The framework is based on current game theory and structural economic models and is intended as a tool for antitrust practice directly targeting consumer welfare analysis. The rest of

(2008) suggests tests based on their theoretical model that links price comparison to welfare effect. However, the tests require homogeneous products and other restrictions.

the paper is organized as follows. Section 2 summarizes potential procompetitive and anticompetitive effects discussed in the literature and in practice. Section 3 provides the main model analyzing exclusive discount with bundled products. Representative variations of the main model are given in Section 4. Section 5 provides numerical examples applying the main model and its variations. Section 6 concludes the paper.

4.2 Antitrust Issues with Exclusive Discount

Various pro- and anticompetitive effects of exclusive discount are explored in the literature and in antitrust practice. Different effects are crucial in different cases and some effects might be irrelevant in certain cases. Important potential anticompetitive effects include foreclosure, price discrimination, discouraged discounting, and reduced product diversity. On the other hand, exclusive discount might also benefit consumers through true discount,[6] preventing free-riding, generating efficiencies, and reducing uncertainties. The framework introduced in the next section captures the net effect of an exclusive discount in the absence of benefits from all the procompetitive effects other than the pricing mechanism itself. In practice, for a given fixed- or setup-cost level, equilibrium profits solved in the framework can be used to predict exit and entry decisions (the foreclosure effect). Influences of an exclusive discount on incentives for discounting or product diversity can be evaluated by applying the framework to different market structures and comparing the equilibriums as shown in Section 4. Potential benefits related to free-riding, efficiencies, and uncertainties are not captured in the framework. Modeling them should be industry specific and data dependent. Thus, when any of these potential benefits are important, conclusions from the framework alone would be conservative in validating an exclusive discount contract.

To highlight different channels through which an exclusive discount might affect consumer welfare, pro- and anticompetitive effects are listed in this section item by item. However, it is worth noting that these effects might be interdependent. For example, exclusive discount may reduce uncertainty and promote investment, which is generally regarded as a procompetitive effect. However, investment promotion is not necessarily procompetitive as

[6] True discount is defined as the lowering of price in the absence of any other procompetitive effects compared to the but-for price without exclusive discount. It emphasizes the impact of the pricing mechanism and is, strictly speaking, a net effect. See more discussion on this in Section 2.2.1.

it might facilitate foreclosure.[7] This implies that we should not evaluate effects separately and then summarize them. Instead, an empirical model targeting net consumer welfare and incorporating important relevant effects would be more appropriate.

4.2.1 Potential Anticompetitive Effects

In this subsection, I briefly describe three common anticompetitive concerns raised in the literature and in practice.[8] In addition, the last part stresses consideration of product diversity, which is rarely considered but could be important. To highlight the effects, the discussion focuses on the chain of causalities of an anticompetitive effect while ignoring interactions with other effects. The same rule applies to the section on procompetitive effects. Instead of matching the conditions for each effect, the empirical setting should be based on industry specifics and the estimated net effect of bundled or exclusive discount is a balance of all relevant anticompetitive and procompetitive effects.

4.2.1.1 Foreclosure

The major antitrust concern on an exclusive discount is that it may foreclose (potential) competitors by depriving them of (part of) the market demand, thereby reducing competition constraints.[9] With reduced individual demand, rival firms may be effectively prevented from reaching efficient scale, discouraged from making investment,[10] and forced to adjust their target market. When the foreclosed market is large enough, the exclusive contract may induce exit or deter entry. The foreclosure effect of an exclusive discount depends on many factors[11] such as the degree of foreclosure, the length of the contract, the importance of the market to the rival suppliers, and availability of potential buyers to which a rival supplier

[7] See Fumagalli et al. (2009) for further discussion.

[8] There are also other less commonly considered anticompetitive effects. For example, bundled discount can be used to internalize pricing externalities in the presence of complementary goods.

[9] There is a large body of literature on the foreclosure effect of exclusive dealing. See, for example, Fumagalli and Motta (2006), Wright (2009), Abito and Wright (2008), Chen and Riordan (2007), Nalebuff (2004, 2005).

[10] Foreclosure discourages investment because the residual market is not large enough to cover the investment. Foreclosure is also likely to stimulate investment for a rival in order to improve its market condition. However, such a story must explain why the rival would not make the same or even greater investment without the pressure of exclusive contract.

[11] See Tom et al. (2000) for conditions and antitrust practice for the foreclosure effect of market-share discounts.

can turn. In practice, price–cost type tests are commonly used and the major issue lies in determining the but-for prices, which should provide predictions about firm strategies under counterfactual settings. The counterfactual analysis demands a structural economic model of consumer preference and other factors that are reasonably linked between the reality and counterfactual worlds.

4.2.1.2 Market Power and Price Discrimination

Suppliers' use of exclusive discounts in the distribution of their products has also been attacked as unlawful price discrimination under the Robinson–Patman Act.[12] For example, the FTC challenged payments by McCormick in exchange for near exclusive shelf space allocations as secondary-line price discrimination under this act.[13] Theoretically, bundled discounts can be used to price discriminate when consumers have heterogeneous preferences. For imperfect competition in the bundled market, mixed bundling allows a supplier to separate the consumers who demand the entire bundle from those who are only interested in a single market. The price-discrimination effect suggests that bundled exclusive discount may hurt consumers even without foreclosing any competitors.[14] However, as the numerical examples in Section 5 imply, the price-discrimination effect might be weak and dominated by procompetitive effects when suppliers also compete in offering the bundled discount.[15]

4.2.1.3 Discourage Discounting

Elhauge (2009a, 2009b) argues that exclusive discounts can perversely discourage discounting when suppliers commit in discount rate or magnitude. The idea is that when a supplier commits in discount rate or magnitude, it has less incentive to compete for free buyers. With the commitment, any price reduction to win sales to free buyers will also lower prices to loyal buyers. This restriction in cutting prices in turn reduces the incentive of its ri-

[12] Building a theoretical model with homogeneous products, Greenlee et al. (2008) proposed tests for the impact of price discrimination from exclusive discount on consumer welfare.

[13] For further discussion, see Kobayashi (2005b).

[14] It is well known that there are simplified theoretical examples where second -degree price discrimination might benefit consumers even with a monopoly. However, it is not clear whether that is also true in the case of exclusive discount without further examinations as in Section 5.

[15] As discussed in Anderson and Leruth (1993), when bundled discount is employed by all suppliers, they compete on many fronts. Thus, competition is intensified, which lowers supplier profits and increases consumer welfare.

vals to cut prices. In addition, bundled exclusive discount may also discourage rivals from discounting if the rivals do not compete in all products in the bundle. In this case, the multimarket supplier can offer a small discount on all products in its bundle to make the bundle more attractive than a large discount of a single product supplier. However, this argument may rely on monopoly power in one of the linked markets. If the supplier faces competitors in each market, then even if the rival on each product market is not sufficient to exert competitive constraint, the rivals together may effectively "offer" a competitive bundle. In addition, when market structure changes, the optimal discount rate or magnitude might be smaller than the committed level, then the distortion in the supplier's strategy might force it to discount more than its optimal level and could instead benefit consumers.

4.2.1.4 Reducing Product Diversity

For consumers with heterogeneous preferences, reduction in product diversity due to an exclusive contract should also be treated as an anticompetitive effect. Note that even partial requirements, such as a market-share discount, reduce product diversity. To satisfy the exclusive requirement, a distributor needs to restrict availability of rivals' products for certain consumers or for certain periods. This provides a crucial implication for tests that concentrate solely on prices: When diversity is important to consumers, showing that consumers enjoy prices lower than the but-for case is insufficient. The consumers might benefit in prices while losing in product diversity. Thus, an economic model taking into account product diversity, such as the discrete-choice model, might be more appropriate.

4.2.2 Potential Procompetitive Effects

Other than using discounts as a tool for more aggressive competition (reduced -price effect), all the procompetitive effects discussed below are related to the impact of the exclusive requirement from the literature of exclusive dealing.[16] These effects are generally hard to verify, let alone to quantify. As a first step in building an empirical framework for exclu-

[16] There are other relatively rare procompetitive arguments. For example, Kolay et al. (2004) shows that all-units discounts can be used to efficiently address double-marginalization problems in the presence of bilateral monopoly. For bundled discount, Greenlee et al. (2008) provides an interesting example with pure bundling where bundling facilitates entry. Kobayashi (2005b) discusses that when demands for products are correlated, a manufacturer may also use bundled discount as a way to facilitate entry into a new market.

sive discount, the empirical model ignores the procompetitive effects other than reducing prices. Thus, the framework will underestimate benefits of the exclusive discount when the procompetitive effects discussed below (other than the reduced-price effect) are important.

4.2.2.1 Reduced Price

Although the literature stresses and examines situations where the discounts are not truly lowering prices, the most natural procompetitive effect of exclusive discount should be the possibility of reducing prices. For bundled discount, exclusive discount might be used to price more aggressively for certain bundles. It is more likely that exclusive discount will strengthen competition when the competitors, instead of being foreclosed from the market, offer exclusive discounts themselves. Even when foreclosure happens, as long as only part of the market is foreclosed and any competitor is still able to reach its minimum efficient scale, it is still possible for exclusive discount to benefit consumers. Klein and Murphy (2008) provides one interesting story on this. In essence, exclusive dealing in the Klein–Murphy model intensifies competition and lowers prices because it replaces individual preferences for different brands with aggregate indifference among brands.

4.2.2.2 Prevention of Distributor Free-Riding

Distributor free-riding occurs when distributors use either manufacturer-supplied promotional investments or manufacturer paid-for promotional efforts to sell rival products.[17] The reason is that manufacturers and distributors have divergent incentives in making an additional sale of the manufacturer's product. When a retailer distributes products of competing manufacturers, a manufacturer benefits from incremental sales of its product from promotion. However, the distributor benefits less because she must take into account the negative externality of the incremental sales of the promoted product on sales of other competing products. Utilizing promotional effort from a particular manufacturer to improve sales of all manufacturers' products is thus profitable for the distributor but detrimental to the manufacturer.

[17] See further discussion on conditions and rationales for exclusive discount in preventing free-riding in Klein and Lerner (2007) and Klein and Murphy (2008).

The potential of free-riding discourages manufacturers from making promotional efforts, thus reducing competition in promotional investment. When promotional efforts facilitate the purchases of the final consumers or help them in making decisions by providing necessary information, the negative impact of free-riding could directly hurt final consumers.

Exclusive dealing solves the free-riding problem by eliminating the distributors' option of selling competing products. Even for incomplete exclusionary contracts like a market-share contract, it still helps by reducing the distributor's benefit and incentive for selling the rival's products.

4.2.2.3 Generating Efficiencies

The exclusive dealing requirement of the discount, either involving bundling or not, has many potential efficiencies. First, it could reduce transaction cost. Second, if the contract results in a higher sale compared to the but-for case, then it recovers fixed cost faster and generates efficiencies through economies of scale. Third, for distributors, exclusive dealing may help to assure quality or assure uniformity. It could also help distributors achieve logistical efficiencies as they deal with fewer suppliers.[18] Fourth, for contracts that involve bundling, bundling also may reduce the transaction and information costs involved in purchasing, distributing, and selling goods and services. If the bundled products are demand-side complements, it might benefit consumers by providing package- or market-integrated products. For example, as suggested by Kobayashi (2005b), "transaction cost savings can explain the use of standardized option packages for automobiles, computer hardware and software, and the packaging of cold remedies and analgesics."

4.2.2.4 Reduced Uncertainty

By establishing an exclusive contract, both the manufacturers and the distributors might benefit from reduced uncertainty. Some benefits from reduced uncertainty are directly cost related, while some are not. Specifying price in the contract protects both parties against price fluctuations. By making demand and supply more predictable, the contract may assure a dependable source of supply, prevent the hold-up problem, and promote investment in

[18] See further discussion in Steuer (2000).

new supply and product enhancements.[19] It is easy to see that the benefit from uncertainty for a specific case will depend on both the degree of uncertainty and its importance. For example, if the relevant market has a record of maintaining a steady price level, then protecting against price fluctuation is hardly a concern in introducing the exclusionary requirement.

4.3 An Empirical Framework for Exclusive Discount

4.3.1 Setup and Demand Function

Consider a static model where suppliers produce heterogeneous products in several related markets. For example, in the pediatric vaccine market, firms produces several different types of vaccines that may or may not be purchased together by consumers. Firms compete in price and maximize their current profits. For the rest of this section, I elaborate with a three-supplier, two-market model that captures key features of bundled and exclusive discount. The model is easily revised to allow more markets and suppliers.

In markets A and B, there are three suppliers. Supplier S_1 sells product A_1 on market A and product B_1 on market B. Supplier S_2 sells product A_2 on market A only. Supplier S_3 sells product B_2 on market B only. In this setting, Supplier S_1 embodies the supplier who might use its advantage in markets A and B to charge different prices conditional on consumers' bundle choices, which may (or may not) hurt other competitors and the consumers. One could also choose to allow all suppliers to compete in all markets and all offer bundled discount as in Section 4.1.2. I leave that as a variation because the literature and antitrust practice suggest that industries where all suppliers can offer bundled discount raise fewer antitrust concerns, especially the concern of the anticompetitive foreclosure effect.

Consumers choose to buy either one product from each market, or one product from market A or B only, or make no purchases at all. Thus, the full set choices for a consumer is:

$$AB = \{A_1, A_2, B_1, B_2, (A_1, B_1), (A_1, B_2), (A_2, B_1), (A_2, B_2), \text{NoPurchase}\}.$$

[19] See further discussion in Gilbert (2000) and Steuer (2000).

Defining an outside good $A_0\,(B_0)$ on market $A\,(B)$, the choice set can be otherwise written as

$$AB = \{(a,b) \ \mid a \in \{A_1, A_2, A_0\}, b \in \{B_1, B_2, B_0\}\}. \tag{1}$$

The utility of a final consumer i choosing a bundle $(a,b) \in AB$, namely product a from market A and product b from market B, is

$$u_{abi} = f(a,b) + \varepsilon_{abi}.$$

When there is no need to specify product a and b, I use n, ab, and (a,b) interchangeably to simplify expressions and stress that choice is over the bundle rather than a single product. As commonly employed in the literature, I further assume that utility is linear in characteristics and prices of product a and b:[20]

$$u_{abi} = X_a\beta_{ai} + X_b\beta_{bi} + \xi_{ab} - \alpha_i\,(p_a + p_b) + \varepsilon_{abi}.$$

Note that we can only identify an unobserved quality for the bundle, ξ_{ab}. As some bundles share common products in market A or B, a logit model assuming i.i.d. ε_{abi} is inappropriate. Specific to industry characteristics, one can use a nested logit model, a general nested logit model[21] or a random-coefficient model. In this example, I illustrate with a nested logit model where products are assigned to four groups: no purchase, purchase in both markets, purchase in market A only, and purchase in market B only:

$$\begin{aligned}
G_0 &: (A_0, B_0) \\
G_1 &: (A_1, B_1), (A_1, B_2), (A_2, B_1), (A_2, B_2) \\
G_2 &: (A_1, B_0), (A_2, B_0) \\
G_3 &: (A_0, B_1), (A_0, B_2).
\end{aligned}$$

[20] The model also permits a nonlinear utility function. I choose a linear function because it is most commonly used in practice.

[21] A cross-nested logit model is more appropriate. For example, all firms might sell in all markets with no bundles consisting of products from different firms. Then the alternatives can be grouped by markets covered as well as by producers. See Pereira et al. (2012) for an example, and see Bierlaire (2006) for how to apply the cross-nested logit model.

The utility function then can be written as

$$\begin{aligned} u_{abi} &= X_a\beta_a + X_b\beta_b + \xi_{ab} - \alpha\left(p_a + p_b\right) + \varepsilon_{abi} \\ &= v_{ab} - \alpha\left(p_a + p_b\right) + \varepsilon_{abi} \\ &= \delta_{ab} + \varepsilon_{abi}, \end{aligned}$$

where for choice alternative n of the total N alternatives and group k of the total K groups for consumer i, the cumulative distribution of $\varepsilon_i = (\varepsilon_{i1,\ldots,}\varepsilon_{iN})$ is

$$F\left(\varepsilon_{i1,\ldots,}\varepsilon_{iN}\right) = \exp\left(-\sum_{k=1}^{K}\left(\sum_{n\in G_k} e^{-\varepsilon_{in}/\lambda_k}\right)^{\lambda_k}\right).$$

λ_k is the parameter that measures the degree of independence in unobserved utility among the alternatives in group k. The mean utility of (A_0, B_0) can be normalized to be 0: $\delta_{00} = 0$. For alternative n in group G_k, choice probability or market share s_n under the nested logit model is

$$s_n = \frac{e^{\delta_n/\lambda_k}\left(\sum_{m\in G_k} e^{\delta_m/\lambda_k}\right)^{\lambda_k - 1}}{\sum_{l=1}^{K}\left(\sum_{m\in G_l} e^{\delta_m/\lambda_l}\right)^{\lambda_l}}. \tag{2}$$

Note that s_n is a function of all prices as it is a function of all mean utilities ($\delta_n\ \forall\ n$).

Once choice alternatives are defined as bundles as in Expression (1), demand function parameters can be estimated using price, market share, product characteristics, and instrument data as in Berry et al. (1995). In cases where individual purchase data are observed,[22] maximum likelihood, control function, or other estimation methods can be applied using Expression (2). See Train (2003) for a thorough discussion.

4.3.2 Supply Side and Pricing Mechanisms

For Supplier S_1, the cost function is $C_1\left(q_{A_1}, q_{B_1}\right)$ with $\frac{\partial C_1}{\partial q_{A_1}} > 0$ and $\frac{\partial C_1}{\partial q_{B_1}} > 0$. For simplicity, let

$$C_1\left(q_{A_1}, q_{B_1}\right) = c_{A_1}q_{A_1} + c_{B_1}q_{B_1} + F_1.$$

[22] One potential example is the pediatric vaccine market, where vaccine choices may be obtained through claim data.

For Suppliers S_2 and S_3, let

$$
\begin{aligned}
C_2\left(q_{A_1}\right) &= c_{A_2}q_{A_2}+F_2; \\
C_3\left(q_{B_2}\right) &= c_{B_2}q_{B_2}+F_3.
\end{aligned}
$$

Fixed costs are omitted in the rest of the discussion as they do not affect market equilibria in the model.

The suppliers' profit functions depend on their pricing mechanisms. In this section, I consider three settings based on strategies allowed for supplier S_1: (i) no discount; (ii) bundled discount; (iii) (perfect) exclusive discount. In applications, pricing mechanisms allowed for all suppliers should be case specific. For example, if single-product suppliers, such as S_2, are found to offer exclusive discount, this can easily be incorporated into the framework in this paper. I allow suppliers S_2 and S_3 to employ exclusive discount in the numerical examples (case (iv) all-product exclusive discount) but omit the formulae in this section for simplicity. Fixed-difference discounts are discussed in Section 4.2.

The prices of Supplier S_1 affect five alternatives:

$$
\left(A_1,B_1\right),\left(A_1,B_2\right),\left(A_2,B_1\right),\left(A_1,B_0\right),\left(A_0,B_1\right).
$$

Let the price when purchasing both products from Supplier S_1 be $d_{A_1B_1}$, the prices when purchasing one product from Supplier S_1 and one from the other suppliers be (p_{A_1},p_{B_1}), the prices when purchasing only in market A or B be (d_{A_1},d_{B_1}). The profit function of Supplier S_1 can be written as

$$
\begin{aligned}
& \pi_1\left(d_{A_1B_1},p_{A_1},p_{B_1},d_{A_1},d_{B_1}\right) \qquad (3)\\
= \; & \left(d_{A_1B_1}-c_{A_1}-c_{B_1}\right)q_{A_1B_1}+\left(p_{A_1}-c_{A_1}\right)q_{A_1B_2} \\
& +\left(p_{B_1}-c_{B_1}\right)q_{A_2B_1}+\left(d_{A_1}-c_{A_1}\right)q_{A_1B_0}+\left(d_{B_1}-c_{B_1}\right)q_{A_0B_1} \\
= \; & M\begin{bmatrix} \left(d_{A_1B_1}-c_{A_1}-c_{B_1}\right)s_{A_1B_1}+\left(p_{A_1}-c_{A_1}\right)s_{A_1B_2} \\ +\left(p_{B_1}-c_{B_1}\right)s_{A_2B_1}+\left(d_{A_1}-c_{A_1}\right)s_{A_1B_0}+\left(d_{B_1}-c_{B_1}\right)s_{A_0B_1}\end{bmatrix}
\end{aligned}
$$

where M is the market size or the total number of final consumers. Note that M does not affect price choices as I assume suppliers can always find an interior solution for optimal prices that generates positive profit.

For (i), no discount, Supplier S_1 sets a common price (p_{A_1}, p_{B_1}) for all alternatives. Namely,

$$d_{A_1B_1} \equiv p_{A_1} + p_{B_1};$$

$$d_{A_1} \equiv p_{A_1};\ d_{B_1} \equiv p_{B_1}.$$

Then the first-order conditions (FOCs) for Supplier S_1 in this case are

$$\begin{aligned}
&s_{A_1B_1} + s_{A_1B_2} + s_{A_1B_0} + (p_{A_1} - c_{A_1})\left(\frac{\partial s_{A_1B_1}}{\partial p_{A_1}} + \frac{\partial s_{A_1B_2}}{\partial p_{A_1}} + \frac{\partial s_{A_1B_0}}{\partial p_{A_1}}\right)\\
&+ (p_{B_1} - c_{B_1})\left(\frac{\partial s_{A_1B_1}}{\partial p_{A_1}} + \frac{\partial s_{A_2B_1}}{\partial p_{A_1}} + \frac{\partial s_{A_0B_1}}{\partial p_{A_1}}\right) = 0\\
&s_{A_1B_1} + s_{A_2B_1} + s_{A_0B_1} + (p_{A_1} - c_{A_1})\left(\frac{\partial s_{A_1B_1}}{\partial p_{B_1}} + \frac{\partial s_{A_1B_2}}{\partial p_{B_1}} + \frac{\partial s_{A_1B_0}}{\partial p_{B_1}}\right)\\
&+ (p_{B_1} - c_{B_1})\left(\frac{\partial s_{A_1B_1}}{\partial p_{B_1}} + \frac{\partial s_{A_2B_1}}{\partial p_{B_1}} + \frac{\partial s_{A_0B_1}}{\partial p_{B_1}}\right) = 0.
\end{aligned} \tag{4}$$

For (ii), bundled discount, discount is only applied when consumers purchase the bundle (A_1, B_1). Stand-alone prices are the same whether consumers purchases products from suppliers S_2 and S_3 or not:

$$d_{A_1} \equiv p_{A_1};\ d_{B_1} \equiv p_{B_1}.$$

Supplier S_1 sets the bundle price and stand-alone prices $(d_{A_1B_1}, p_{A_1}, p_{B_1})$ and the FOCs are

$$\begin{aligned}
&s_{A_1B_1} + (d_{A_1B_1} - c_{A_1} - c_{B_1})\frac{\partial s_{A_1B_1}}{\partial d_{A_1B_1}} + (p_{A_1} - c_{A_1})\left(\frac{\partial s_{A_1B_2}}{\partial d_{A_1B_1}} + \frac{\partial s_{A_1B_0}}{\partial d_{A_1B_1}}\right)\\
&+ (p_{B_1} - c_{B_1})\left(\frac{\partial s_{A_2B_1}}{\partial d_{A_1B_1}} + \frac{\partial s_{A_0B_1}}{\partial d_{A_1B_1}}\right) = 0\\
&s_{A_1B_2} + s_{A_1B_0} + (d_{A_1B_1} - c_{A_1} - c_{B_1})\frac{\partial s_{A_1B_1}}{\partial p_{A_1}} + (p_{A_1} - c_{A_1})\left(\frac{\partial s_{A_1B_2}}{\partial p_{A_1}} + \frac{\partial s_{A_1B_0}}{\partial p_{A_1}}\right)\\
&+ (p_{B_1} - c_{B_1})\left(\frac{\partial s_{A_2B_1}}{\partial p_{A_1}} + \frac{\partial s_{A_0B_1}}{\partial p_{A_1}}\right) = 0\\
&s_{A_2B_1} + s_{A_0B_1} + (d_{A_1B_1} - c_{A_1} - c_{B_1})\frac{\partial s_{A_1B_1}}{\partial p_{B_1}} + (p_{A_1} - c_{A_1})\left(\frac{\partial s_{A_1B_2}}{\partial p_{B_1}} + \frac{\partial s_{A_1B_0}}{\partial p_{B_1}}\right)\\
&+ (p_{B_1} - c_{B_1})\left(\frac{\partial s_{A_2B_1}}{\partial p_{B_1}} + \frac{\partial s_{A_0B_1}}{\partial p_{B_1}}\right) = 0.
\end{aligned} \tag{5}$$

For (iii), exclusive discount, consumers receive the discount if and only if they do not purchase from other suppliers. I assume that there is no further discount from purchasing the bundle (A_1, B_1).[23] Then Supplier S_1 sets $(d_{A_1}, d_{B_1}, p_{A_1}, p_{B_1})$ and

$$d_{A_1B_1} \equiv d_{A_1} + d_{B_1}.$$

The FOCs are

$$\begin{aligned}
&s_{A_1B_1} + s_{A_1B_0} + (d_{A_1} - c_{A_1})\left(\tfrac{\partial s_{A_1B_1}}{\partial d_{A_1}} + \tfrac{\partial s_{A_1B_0}}{\partial d_{A_1}}\right) + (d_{B_1} - a_{B_1})\left(\tfrac{\partial s_{A_1B_1}}{\partial d_{A_1}} + \tfrac{\partial s_{A_0B_1}}{\partial d_{A_1}}\right)\\
&+ (p_{A_1} - c_{A_1})\tfrac{\partial s_{A_1B_2}}{\partial d_{A_1}} + (p_{B_1} - c_{B_1})\tfrac{\partial s_{A_2B_1}}{\partial d_{A_1}} = 0\\
&s_{A_1B_1} + s_{A_0B_1} + (d_{A_1} - c_{A_1})\left(\tfrac{\partial s_{A_1B_1}}{\partial d_{B_1}} + \tfrac{\partial s_{A_1B_0}}{\partial d_{B_1}}\right) + (d_{B_1} - a_{B_1})\left(\tfrac{\partial s_{A_1B_1}}{\partial d_{B_1}} + \tfrac{\partial s_{A_0B_1}}{\partial d_{B_1}}\right)\\
&+ (p_{A_1} - c_{A_1})\tfrac{\partial s_{A_1B_2}}{\partial d_{B_1}} + (p_{B_1} - c_{B_1})\tfrac{\partial s_{A_2B_1}}{\partial d_{B_1}} = 0\\
&s_{A_1B_2} + (d_{A_1} - c_{A_1})\left(\tfrac{\partial s_{A_1B_1}}{\partial p_{A_1}} + \tfrac{\partial s_{A_1B_0}}{\partial p_{A_1}}\right) + (d_{B_1} - a_{B_1})\left(\tfrac{\partial s_{A_1B_1}}{\partial p_{A_1}} + \tfrac{\partial s_{A_0B_1}}{\partial p_{A_1}}\right)\\
&+ (p_{A_1} - c_{A_1})\tfrac{\partial s_{A_1B_2}}{\partial p_{A_1}} + (p_{B_1} - c_{B_1})\tfrac{\partial s_{A_2B_1}}{\partial p_{A_1}} = 0\\
&s_{A_2B_1} + (d_{A_1} - c_{A_1})\left(\tfrac{\partial s_{A_1B_1}}{\partial p_{B_1}} + \tfrac{\partial s_{A_1B_0}}{\partial p_{B_1}}\right) + (d_{B_1} - a_{B_1})\left(\tfrac{\partial s_{A_1B_1}}{\partial p_{B_1}} + \tfrac{\partial s_{A_0B_1}}{\partial p_{B_1}}\right)\\
&+ (p_{A_1} - c_{A_1})\tfrac{\partial s_{A_1B_2}}{\partial p_{B_1}} + (p_{B_1} - c_{B_1})\tfrac{\partial s_{A_2B_1}}{\partial p_{B_1}} = 0.
\end{aligned} \tag{6}$$

The FOCs highlight the difference between bundled discount and exclusive discount. In the numerical example in Section 5, I show that the two types of discount provide different incentives for the suppliers and result in different prices and consumer surpluses.

Supplier S_2 chooses p_{A_2} to maximize its profit:

$$\begin{aligned}
\pi_2(p_{A_2}) &= (p_{A_2} - c_{A_2})(q_{A_2B_1} + q_{A_2B_2} + q_{A_2B_0}) \\
&= M(p_{A_2} - c_{A_2})(s_{A_2B_1} + s_{A_2B_2} + s_{A_2B_0}).
\end{aligned} \tag{7}$$

FOC:

$$s_{A_2B_1} + s_{A_2B_2} + s_{A_2B_0} + (p_{A_2} - c_{A_2})\left(\frac{\partial s_{A_2B_1}}{\partial p_{A_2}} + \frac{\partial s_{A_2B_2}}{\partial p_{A_2}} + \frac{\partial s_{A_2B_0}}{\partial p_{A_2}}\right) = 0.$$

[23] Specific to industry features, one can also allow additional discount when purchasing the bundle (A_1, B_1). Then Supplier S_1 sets $(d_{A_1B_1}, d_{A_1}, d_{B_1}, p_{A_1}, p_{B_1})$.

Supplier S_3 chooses p_{B_2} to maximize its profit:

$$\begin{aligned}\pi_3\left(p_{B_2}\right) &= \left(p_{B_2}-c_{B_2}\right)\left(q_{A_1B_2}+q_{A_2B_2}+q_{A_0B_2}\right) \\ &= M\left(p_{B_2}-c_{B_2}\right)\left(s_{A_1B_2}+s_{A_2B_2}+s_{A_0B_2}\right).\end{aligned} \tag{8}$$

FOC:

$$s_{A_1B_2}+s_{A_2B_2}+s_{A_0B_2}+\left(p_{B_2}-c_{B_2}\right)\left(\frac{\partial s_{A_1B_2}}{\partial p_{B_2}}+\frac{\partial s_{A_2B_2}}{\partial p_{B_2}}+\frac{\partial s_{A_0B_2}}{\partial p_{B_2}}\right)=0.$$

I assume that all suppliers choose their prices simultaneously. The FOCs of the profit-maximization problems (3), (7), and (8) give the suppliers' best replies. Solving them together provides the equilibrium prices.[24] FOCs for this bundle alternative model are more complicated than the single-product alternative case. Important steps in solving the FOCs are given in the appendix.

Note that when two firms set two prices that together determine the total price of a cross-firm alternative and the two prices are not part of the total prices of any other alternatives, one possible equilibrium is that the two prices are set effectively to infinity so that market share of the alternative is effectively 0.[25] This is part of an equilibrium because a unilateral price drop will not affect the alternative's market share.

4.3.3 More on the Cost Function

Policy evaluation requires computing equilibrium strategies—in this case, prices for the counterfactual settings using the FOCs. If there is no cost data, after the demand function is estimated based on expression (2), marginal costs can be solved using FOCs for observed prices. These marginal costs solved from reality cases are then used in solving prices in

[24] As in Berry et al. (1995), existence of an equilibrium is assumed and can be checked numerically. Besides, the FOCs only provide necessary conditions, the solutions are numerically checked to be profit maximizers.

[25] Mathematically, the market share is never 0 unless the price goes to infinity. Numerically, a high enough price will give a 0 market share, which is what "effectively infinity" and "effectively 0" mean. Given the price of one firm, if the price of the other firm must be below marginal cost to generate a positive market share for the alternative, then any price above marginal cost that keeps the market share effectively 0 is a best response. Thus, I also find equilibria where only one of the two prices for the alternative is set to effectively 0 in numerical results.

the counterfactual settings. FOCs in the previous section are suitable for solving marginal costs when the observed reality case is (i) no discount. In this case, the total number of price variables equals the total number of products and the total number of marginal cost parameters is as in the FOCs (4). Note that the FOCs are linear in the marginal costs. Thus, as long as the coefficient matrix is invertible, the marginal costs can be uniquely determined using the FOCs.

However, when the observed reality case is instead (ii), bundled discount, or (iii), exclusive discount, FOCs (5) and (6) cannot be directly used to back out the marginal cost parameters. This is because the number of price variables is different depending on the discount mechanism. For cases (ii) and (iii), the number of price variables is larger than number of marginal cost parameters. Thus, if the augmented coefficient matrix for the FOCs when treating marginal cost as the unknowns has full rank, then there is no solution for the marginal costs. Empirically, the solution is simply to allow more cost parameters so that number of FOCs equals number of cost parameters. For example, for (ii), bundled discount, one can allow the sum of marginal cost in offering A_1 and B_1 together as a new parameter $a_{A_1B_1}$. Then the FOCs for Supplier S_1 become

$$\begin{aligned}
&s_{A_1B_1} + (d_{A_1B_1} - a_{A_1B_1})\frac{\partial s_{A_1B_1}}{\partial d_{A_1B_1}} + (p_{A_1} - c_{A_1})\left(\frac{\partial s_{A_1B_2}}{\partial d_{A_1B_1}} + \frac{\partial s_{A_1B_0}}{\partial d_{A_1B_1}}\right)\\
&+(p_{B_1} - c_{B_1})\left(\frac{\partial s_{A_2B_1}}{\partial d_{A_1B_1}} + \frac{\partial s_{A_0B_1}}{\partial d_{A_1B_1}}\right) = 0\\
&s_{A_1B_2} + s_{A_1B_0} + (d_{A_1B_1} - a_{A_1B_1})\frac{\partial s_{A_1B_1}}{\partial p_{A_1}} + (p_{A_1} - c_{A_1})\left(\frac{\partial s_{A_1B_2}}{\partial p_{A_1}} + \frac{\partial s_{A_1B_0}}{\partial p_{A_1}}\right)\\
&+(p_{B_1} - c_{B_1})\left(\frac{\partial s_{A_2B_1}}{\partial p_{A_1}} + \frac{\partial s_{A_0B_1}}{\partial p_{A_1}}\right) = 0\\
&s_{A_2B_1} + s_{A_0B_1} + (d_{A_1B_1} - a_{A_1B_1})\frac{\partial s_{A_1B_1}}{\partial p_{B_1}} + (p_{A_1} - c_{A_1})\left(\frac{\partial s_{A_1B_2}}{\partial p_{B_1}} + \frac{\partial s_{A_1B_0}}{\partial p_{B_1}}\right)\\
&+(p_{B_1} - c_{B_1})\left(\frac{\partial s_{A_2B_1}}{\partial p_{B_1}} + \frac{\partial s_{A_0B_1}}{\partial p_{B_1}}\right) = 0.
\end{aligned} \tag{9}$$

Similarly, for (iii), exclusive discount, one can allow marginal costs to be different when

A_1 or B_1 are offered in discount as a_{A_1} and a_{B_1}. Then the FOCs for Supplier S_1 become

$$\begin{aligned}
& s_{A_1B_1} + s_{A_1B_0} + (d_{A_1} - a_{A_1})\left(\frac{\partial s_{A_1B_1}}{\partial d_{A_1}} + \frac{\partial s_{A_1B_0}}{\partial d_{A_1}}\right) + (d_{B_1} - a_{B_1})\left(\frac{\partial s_{A_1B_1}}{\partial d_{A_1}} + \frac{\partial s_{A_0B_1}}{\partial d_{A_1}}\right) \\
& + (p_{A_1} - c_{A_1})\frac{\partial s_{A_1B_2}}{\partial d_{A_1}} + (p_{B_1} - c_{B_1})\frac{\partial s_{A_2B_1}}{\partial d_{A_1}} = 0 \\
& s_{A_1B_1} + s_{A_0B_1} + (d_{A_1} - a_{A_1})\left(\frac{\partial s_{A_1B_1}}{\partial d_{B_1}} + \frac{\partial s_{A_1B_0}}{\partial d_{B_1}}\right) + (d_{B_1} - a_{B_1})\left(\frac{\partial s_{A_1B_1}}{\partial d_{B_1}} + \frac{\partial s_{A_0B_1}}{\partial d_{B_1}}\right) \\
& + (p_{A_1} - c_{A_1})\frac{\partial s_{A_1B_2}}{\partial d_{B_1}} + (p_{B_1} - c_{B_1})\frac{\partial s_{A_2B_1}}{\partial d_{B_1}} = 0 \\
& s_{A_1B_2} + (d_{A_1} - a_{A_1})\left(\frac{\partial s_{A_1B_1}}{\partial p_{A_1}} + \frac{\partial s_{A_1B_0}}{\partial p_{A_1}}\right) + (d_{B_1} - a_{B_1})\left(\frac{\partial s_{A_1B_1}}{\partial p_{A_1}} + \frac{\partial s_{A_0B_1}}{\partial p_{A_1}}\right) \\
& + (p_{A_1} - c_{A_1})\frac{\partial s_{A_1B_2}}{\partial p_{A_1}} + (p_{B_1} - c_{B_1})\frac{\partial s_{A_2B_1}}{\partial p_{A_1}} = 0 \\
& s_{A_2B_1} + (d_{A_1} - a_{A_1})\left(\frac{\partial s_{A_1B_1}}{\partial p_{B_1}} + \frac{\partial s_{A_1B_0}}{\partial p_{B_1}}\right) + (d_{B_1} - a_{B_1})\left(\frac{\partial s_{A_1B_1}}{\partial p_{B_1}} + \frac{\partial s_{A_0B_1}}{\partial p_{B_1}}\right) \\
& + (p_{A_1} - c_{A_1})\frac{\partial s_{A_1B_2}}{\partial p_{B_1}} + (p_{B_1} - c_{B_1})\frac{\partial s_{A_2B_1}}{\partial p_{B_1}} = 0.
\end{aligned} \tag{10}$$

Empirically, if we observe more price variables in the data, we can back out a more complicated the cost-function structure. Then the chosen cost-function structure is applied to *all* cases. Note that theoretically there is nothing prevents

$$a_{A_1B_1} = c_{A_1} + c_{B_1}$$

or

$$a_{A_1} = c_{A_1};\ a_{B_1} = c_{B_1}$$

in the model. Thus, theoretically, if the true cost function is a simple special case of the more general cost function specified in the model, then solving cost parameters from the FOCs will result in this special case. Specifically, backing out the cost parameters from the FOCs will tell us whether there is a difference in cost with the discount cases. For bundled discount, one intuition on different marginal costs is that there could be distributional savings in providing A_1 and B_1 together. For exclusive discount, one intuition is based on the fact that there is an additional cost in monitoring whether consumers are violating the exclusive agreement. In addition, whenever a contract is signed for the discount, the monitoring cost may present savings from reduced uncertainties as well. These contract effects are magnified when suppliers instead sell to intermediate distributors who make bulk purchases.

If (ii), bundled discount, is the observed reality case and cost parameters are backed out from FOCs (9), then when applied to counterfactual case (i), no discount, parameter $a_{A_1B_1}$ is not used. When applied to counterfactual case (iii), exclusive discount, one can either not use $a_{A_1B_1}$ and stay with FOCs (6) or, instead, let the cost in the exclusive case to be proportional, for example,

$$a_{A_1} = a_{A_1B_1} \cdot \frac{c_{A_1}}{c_{A_1} + c_{B_1}}.$$

Similarly, if (iii), exclusive discount, is the observed reality case and cost parameters are backed out from FOCs (10), then a_{A_1} and a_{B_1} can be ignored in the counterfactual cases or we can let $a_{A_1B_1} = a_{A_1} + a_{B_1}$. Without additional cost data, the choices should depend on the econometrician's understanding of the industry features.

Note that one can also have nonlinear cost functions as long as the number of cost parameters equals the number of FOCs (assuming that the coefficient matrix from the FOCs always has full rank). When there is additional cost-side data, it can be used to enrich the cost function. One example is marginal cost as a function of product characteristics as in Berry et al. (1995). Again, the econometrician should make sure (a) the cost parameters can be exactly backed out from the reality-case FOCs and (b) the cost parameters can be reasonably applied to the counterfactual cases.

4.3.4 Welfare Effect

After the cost parameters are solved using data from the reality case, they can be used in the FOCs for the counterfactual cases to solve for the counterfactual prices. Prices across cases can then be used to evaluate the impact of the bundled- or exclusive-discount contracts. In addition, because Hicksian and Marshallian demand functions are identical in the nested logit discrete-choice model, consumer surplus, CS, can be obtained simply by integrating the demand function.[26] As pointed out by Nevo (2011), for logit model,

$$CS = \frac{M \ln\left(\frac{1}{s_0}\right)}{\alpha} + C \tag{11}$$

[26] Change in CS is a compensation variation, and CS tends to be overestimated in a logit-based model.

where C is a constant representing that the absolute value of consumer surplus cannot be determined. Following the literature on GEV models since McFadden (1978), it can also be shown that nested and general nested logit models also satisfy Equation (11). Particularly, Equation (11) implies that the formula for consumer surplus with a nested logit model is

$$CS = \frac{M \cdot ln(\sum_k (\sum_{n \in G_k} e^{\frac{\delta_n}{\lambda_k}})^{\lambda_k})}{\alpha} + C.$$

For a general nested logit model, it is

$$CS = \frac{M \cdot ln(\sum_k (\sum_{n \in G_k} \left(\omega_{nk} e^{\delta_n}\right)^{\frac{1}{\lambda_k}})^{\lambda_k})}{\alpha} + C,$$

where $\omega_{nk} \geq 0 \; \forall \; n, k$ captures the portion of alternative n that is allocated to group k and

$$\sum_k \omega_{nk} = 1 \; \forall \; n.$$

For a random coefficient model, it is

$$CS = M \int \frac{\ln\left(\sum_n e^{\delta_{ni}}\right)}{\alpha_i} f\left(\theta_i\right) d\theta_i,$$

where θ_i denotes the random coefficients that vary across consumers.

4.4 Variations of the Framework

4.4.1 Other Market Structures

Theoretically, the model structure works with any number of suppliers covering any number of markets. Here I discuss two representative structures. The first has a monopoly in one market, demonstrated by removing Supplier S_3 in the main model. This variation addresses the issue of extending market power in one market to another through bundled discount. The second has a duopoly in both markets, which merges suppliers S_2 and S_3 and allows the merged new supplier to offer bundled or exclusive discount as well. The model can be used to provide judgments over the argument used by some courts that bundled discount is not anticompetitive if the other firms can also offer it.

4.4.1.1 Monopoly in One Market

Take the setup in Section 3, but assume that there is no Supplier S_3. That is, the data (prices, market shares, and others) observed are from competition between Suppliers S_1 and S_2. The set of all alternatives is

$$AB' = \{(a,b) \mid a \in \{A_1, A_2, A_0\}, b \in \{B_1, B_0\}\},$$

and the alternatives are grouped as

$$\begin{aligned} G_0 &: (A_0, B_0) \\ G_1 &: (A_1, B_1), (A_2, B_1) \\ G_2 &: (A_1, B_0), (A_2, B_0) \\ G_3 &: (A_0, B_1) \end{aligned}$$

In this market structure, Supplier S_1 has a monopoly in market B. A natural question in antitrust practice is whether this monopoly power can be extended to market A through bundled or exclusive discount. As in Nalebuff (2008), monopoly power is extended when (1) in equilibrium the monopoly Supplier S_1 lowers its price in market B for consumers who purchase from it exclusively in market A and (2) this bundling increases profit of Supplier S_1 compared to the no-discount case. For the exclusive-discount case, discount price d_{B_1} is directly observable. For the bundled-discount case, only the total bundled price is observed. Assuming the effective price of A_1 in the bundle is no lower than p_{A_1}, then $d_{A_1B_1} - p_{A_1}$ provides an upper bound of the effective price of B_1 in the bundle. For a given industry, equilibrium prices, profits, and consumer surplus can be compared between the no-discount case and the bundled-discount case. The numerical examples in Section 5 suggest that the conclusion should be case specific. They can also be used to compare this monopoly-in-one-market structure with the main model. The examples suggest that qualitative results do not naturally extend from the main model to the new structure.

4.4.1.2 Duopoly in Both Markets

In this market structure, I assume that all suppliers produce in both markets A and B and can offer bundled or exclusive discount.[27] Particularly, I assume that products A_2 and B_2 are produced by the same firm. Thus, when compared to the main model, the duopoly structure can be used to evaluate a hypothetical merger between Suppliers S_2 and S_3. Another purpose is to incorporate the situations where more than one firm offers bundled or exclusive discount. There are arguments in practice that bundled discount is not anticompetitive if the other firms can also offer it. Although the arguments seems to be reasonable for symmetric firms, it is not obvious whether it is also the case when one firm enjoys a cost or quality advantage. In addition, bundled or exclusive discount in linked markets A and B may be used for price discrimination to segregate markets and reduce competition. The numerical examples in Section 5 explore possible welfare effects under this market structure.

4.4.2 Commitment in Discounts

In Section 3, I assume that there is no fixed relationship between discount prices and regular prices. In practice, suppliers sometimes must make a commitment to lower the discount prices whenever the regular prices are lower to convince consumers to accept the exclusive purchase requirement. The commitment could come in the form of fixed value or percentage discount. For example, for product j,

$$d_j = p_j - \overline{d}_j \tag{12}$$

or

$$d_j = \left(1 - \mu_j\right) p_j. \tag{13}$$

When $\overline{d}_j$ (or μ_j) are choice variables, the optimal solution has no difference from those in Section 3. However, if $\overline{d}_j$ (or μ_j) is fixed when there is an environmental change in

[27] In the corresponding examples in Section 5, I assume that either both suppliers choose bundled discount or both choose exclusive discount. However, one can also let the two suppliers employ different pricing mechanisms if that is the case in an industry.

the markets, commitment in $\overline{d}_j$ (or μ_j) affects the supplier's incentive for competition as argued in Elhauge (2009a).[28] Elhauge (2009a) argues that when a new supplier enters the market, a discount commitment discourages the incumbent to compete for free buyers even when buyers do not commit. However, the result is derived from a stylized theoretical model. Elhauge implicitly assumes that, in a homogeneous product setup, a buyer would not switch suppliers if the incumbent can match the price of the entrant. While this is a reasonable assumption for homogeneous products, it seems to be crucial for the result. In this section, I provide a model to evaluate the commitment effect under heterogeneous product competition as in the main model. A numerical counterpart is given in Section 5. This model and its numerical counterpart can then be used to test the generality of Elhauge's result in the heterogeneous product setup commonly used in empirical work. Note that I use the setup in Section 3 where suppliers compete in more than one market to provide a broad framework for empirical application. The modeling idea readily extends to single-market competition.

I again illustrate the idea with an example, but the model can be generalized to more suppliers with different product sets. Take the setup in the main model. That is, the data (prices, market shares, and others) observed are from competition between Suppliers S_1, S_2, and S_3. Estimate the model and solve for $\left(\overline{d}_{A_1}, \overline{d}_{B_1}\right)$ (or $\left(\mu_{A_1}, \mu_{B_1}\right)$) using Equation (12) (or (13)) for observed prices and discounts of the last year in the data before the market structure changes. Then assume that a hypothetical entrant, Supplier S_4, who produces product A_3 enters the market. The new market equilibrium can be solved with $\left(\overline{d}_{A_1}, \overline{d}_{B_1}\right)$ (or $\left(\mu_{A_1}, \mu_{B_1}\right)$) fixed and (p_{A_1}, p_{B_1}) as the only choice variables for Supplier S_1. As Supplier S_4 is a hypothetical entrant, no cost parameter can be backed out from the data. For the purpose of antitrust analysis, allow product A_3 to have the same cost and quality (mean utility) as product A_1. As a benchmark case, the market equilibrium without the constraint of fixed $\left(\overline{d}_{A_1}, \overline{d}_{B_1}\right)$ (or $\left(\mu_{A_1}, \mu_{B_1}\right)$) can also be calculated. Comparing equilibrium prices with consumer surplus evaluates the impact of commitment in discounts. Note that consumers are assumed to have the freedom to switch suppliers when Supplier S_4 enters as in Elhauge

[28] Elhauge labeled it "loyalty discount." To avoid confusion with loyalty discount for buyers making repeated purchases from the seller, I will refer to it as fixed difference discount or discount with commitment.

(2009a). The results of the numerical analysis are presented in the appendix.

4.5 Numerical Examples

In the numerical examples, I assume the demand function is estimated from some dataset via a nested logit model with the market-share function given as in Expression (2).[29] For the cost function, I assume that the data is obtained from the no-discount case and the marginal cost is backed out from the FOCs (4). The fixed cost is set to zero for all products because endogenous entry and exit are not allowed in the model. The equilibrium profits then provide the fixed-cost thresholds for firms to be active.

With demand estimation, parameters in the model can be simplified. Specifically, as product characteristics are given,

$$v_{ij} \equiv X_{ij}\beta + \xi_{ij}$$

is a fixed number for each alternative (A_i, B_j). The utility function can be written as

$$u_{ij} = \delta_{ij} + \varepsilon_{ij} = v_{ij} - \alpha p_{ij} + \varepsilon_{ij},$$

where v_{ij} and α are given based on data and demand estimation. In this section, I evaluate the market equilibrium for several different settings of the parameters: consumer preference over price α, within-group correlation parameter λ_k, marginal costs (c_i, c_j), and mean price-excluded utilities v_{ij}. The purpose is to illustrate possible results from the model. Welfare effects are found to be qualitatively different across settings, suggesting the effect of bundled or exclusive discount is case specific.

For each parameter setting, I compare equilibrium price, market share, profit, and consumer surplus across several different market structures: (a) the main model described in Section 3 where a multimarket supplier competes against a single-market rival on each market; (b) monopoly in one market described in Section 4.1.1 where the multimarket supplier faces no competition in market B; and (c) duopoly in both markets described in Section

[29] As demand estimation is standard in the literature once alternatives are defined as bundles, I focus on computing counterfactuals and evaluating consumer-welfare effects in the numerical examples.

4.1.2 where two multimarket suppliers compete in both markets. I consider four pricing mechanisms: (i) no discount, (ii) bundled discount for Supplier S_1, (iii) exclusive discount for Supplier S_1, (iv) all-product exclusive discount.

4.5.1 Setting 1: Quasihomogeneous Products

In this first setting, I assume that product characteristics are identical for all products in markets A and B with mean price-excluded utilities v_{ij} being the same within a group. Specifically, I assume that

$$v_{ij} = \begin{cases} v_{AB} = 4 & \text{if } i > 0, j > 0 \\ v_A = 2 & \text{if } i > 0, j = 0 \\ v_B = 2 & \text{if } i = 0, j > 0. \end{cases}$$

Marginal costs are also assumed to be the same across products as

$$c_{A_i} = c_{B_j} = 1 \;\forall\; i, j > 0.$$

Thus, products are different to consumers only in terms of prices and unobserved individual preferences over alternatives ε_{ij}, hence the name quasihomogeneous products.

Tables 1–4 provide the equilibrium when $\alpha = 1$ and $\lambda_k = \overline{\lambda} = 0.6 \;\forall\; k$. To understand the results here, we need to examine the impact of different price mechanisms at the choice-alternative (bundle) level of competition. It is important to differentiate between within-firm alternatives, where all products in the alternative are produced by the same supplier, from cross-firm alternatives, where product A and B are from different suppliers. (See classification in Table 2.) A supplier enjoys the full benefit of lowering prices for its within-firm alternatives but does not for its cross-firm alternatives. In fact, as within-firm and cross-firm alternatives compete directly against each other, the suppliers have incentives to raise prices for cross-firm alternatives while lowering prices for within-firm alternatives under fierce competition. From the perspective of S_1, within-firm alternatives include $(A_1, B_1), (A_1, B_0), (A_0, B_1)$ while cross-firm alternatives include $(A_1, B_2), (A_2, B_1)$. For (i), the no-discount case, Supplier S_1 cannot set separate prices for within-firm and cross-firm alternatives. Thus, it cannot raise or lower prices of the within-firm alternatives alone. For (ii), the bundled-discount case, Supplier S_1 can set a separate price for the within-firm alternative (A_1, B_1). Its ability to raise the stand-alone prices for cross-firm alternatives is still limited as it will also raise the prices for within-firm alternatives (A_1, B_0) and (A_0, B_1). Looking at it from a different angle, Supplier S_1 also has less incentive to lower the prices for (A_1, B_0) and (A_0, B_1) as that also lowers prices for cross-firm alternatives and erodes profits from (A_1, B_1). For (iii), the exclusive-discount case, Supplier S_1 can price differently for within-firm alternatives and cross-firm alternatives. However, with regard to the total price for alternative (A_1, B_1), Supplier S_1 has less freedom than in the bundled-discount case.

Several observations can be made from this quasihomogeneous example. First, consumer-welfare effects are presented in the last row of Table 1. For all market structures (a)–(c), the consumer surplus under (iii), exclusive discount, is larger than those under (ii), bundled discount, and (i), no discount. For example, in the main model, consumer surplus is 1.92 under exclusive discount, but only 1.88 and 1.87 under no discount and bundled discount, respectively.[30] In addition, bundled discount reduces consumer surplus when only offered by one dominant firm (from 1.88 to 1.87 in market structure (a) and from 1.39 to 1.36 in (b)), while it benefits consumers in the duopoly structure (from 1.88 to 1.98) where both firms offer bundled discount. These qualitative results on consumer surplus can be extended to a wide range of α and λ_k. For example, in the main model, consumer surplus in the bundled-discount case is lower than in the no-discount case while consumer surplus in the exclusive-discount case is higher than in the no-discount case, at least for any $(\alpha, \overline{\lambda}) \in (1 \times \{0.5, 0.51, \ldots, 1\}) \cup (\{1, 1.01, \ldots, 2\} \times 1)$.[31]

Previous discussion on price mechanisms and alternative-level competition provides some insight on these results. Table 2 presents the prices at the alternative level which are con-

[30] The differences in consumer surplus get much larger when demand is less elastic. For example, when $\alpha = 0.5$, consumer surplus under (i), no-discount, is 30.14 while consumer surplus under (ii), bundled-discount, is 25.77. Also note that the effects get larger when competitors are actually foreclosed from the market. For example, if the bundled discount forces Supplier S_3 to exit, then the consumer surplus is 1.88 under no discount and market structure (a), compared to 1.36 under bundled discount and market structure (b).

[31] The result in fact holds for more general cases such as when $\lambda = (0.5, 1, 1)$. The complete results of varying (α, λ) and the results for other models and different (α, λ) settings are available upon request.

structed from product prices in Table 3. Green indicates the lowest price for an alternative among the price mechanisms (i), (ii), and (iii), while red indicates the highest price. Exclusive discount leads to the lowest prices for single-product alternatives (the last four rows involving product A_0 or B_0). This is because Supplier S_1 can set a discount on single-product alternatives without lowering prices for the cross-firm alternatives. This, however, triggers fierce competition, lowers rivals' prices, and benefits the consumers. It also explains why exclusive discount leads to the highest prices for cross-firm alternatives (A_1, B_2) and (A_2, B_1) (at 4.31 in the main model): separate high stand-alone prices can be chosen to shift demand to within-firm alternatives of Supplier S_1. Under bundled discount, even though consumers who prefer (A_1, B_1) can enjoy the lowest price (3.4 for the main model), prices for alternatives (A_1, B_0) and (A_0, B_1) are the highest as they are also part of the prices for cross-firm alternatives. Under this numerical setting, it turns out consumer benefits from the single-product alternatives under exclusive discount dominate while benefits from (A_1, B_1) under bundled discount are not large enough to cover losses from high stand-alone prices, p_{A_1} and p_{B_1}. Lastly, for the duopoly structure, $d_{A_2B_2}$ can also be set separately and provides additional pressure to lower $d_{A_1B_1}$. Under this numerical setting, these additional competition pressures make consumers better off compared with the no -discount case.

The second observation is on profits across different price mechanisms as shown in Table 1. In all market structures, bundled discount generates the highest profit when only offered by Supplier S_1 (first row of market structures (a) and (b)).[32] As all alternatives are substitutes of each other, the ability to separately control the price and exploit the full profit for an alternative, (A_1, B_1), generates higher profits, especially when no competitors possess the same power. In addition, in all market structures, exclusive discount leaves the competitors (Suppliers S_2, S_3, and S_4) with the lowest profits. As discussed before, competition is most intensive under exclusive discount and the competitors are forced to lower prices which lowers profits. Recall that all profits are excluding fixed costs. Hence, with high enough fixed cost, exclusive discount is most powerful in foreclosing competitors (under this parameter setting). Lastly, comparing profits of Suppliers S_1 and S_2 in the

[32] Bundled discount reduces profits in the duopoly structure in this parameter setting. However it could increase profits in other settings. For example, when all parameters are the same but $\overline{\lambda} = 1$.

main model (a) with those in the monopoly structure (b) provides insight on the impact of having monopoly power in market B. The first two rows in Table 1 show that for any price mechanism, removing Supplier S_3 increases the profits of Supplier S_1 while weakly reducing the profits of Supplier S_2. For example, when Supplier S_3 is excluded from the market (going from market structure (a) to (b)), under exclusive discount, the profits of Supplier S_1 increase from 0.60 to 0.88 while the profits of Supplier S_2 decrease from 0.25 to 0.24. However, this is the case even under the no-discount case. Hence Supplier S_1 benefits when having monopoly power in market B, but it is not necessarily through bundled or exclusive discount. In fact, prices on market A and the profits of Supplier S_2 are only slightly affected, indicating that most increase in profit of Supplier S_1 come solely from the lower competition in market B.

Third, all discounts are lower than the but-for prices (price under the no-discount case) and stand-alone prices are higher than the but-for prices. For example, in the main model for product A_1, the exclusive-discount price, stand-alone price, and but-for price are, respectively, 1.84, 1.97, and 2.39. Thus, the actual discounts are not as large as nominal discounts.[33] The lower discount price and higher stand-alone prices help shift demand from cross-firm to within-firm alternatives.

Table 1: Profits and CS of Structures (a), (b), and (c) for Setting 1

Profit	(a) Main Model				(b) Monopoly in B				(c) Duopoly		
Firm	(i)	(ii)	(iii)	(iv)	(i)	(ii)	(iii)	(iv)	(i)	(ii)	(iii)
S_1	0.59	0.64	0.60	0.60	0.86	0.94	0.88	0.84	0.59	0.55	0.48
S_2	0.30	0.27	0.25	0.25	0.30	0.27	0.24	0.19	0.59	0.55	0.48
S_3	0.30	0.27	0.25	0.25	N/A	N/A	N/A	N/A	N/A	N/A	N/A
CS	1.88	1.87	1.92	1.92	1.39	1.36	1.40	1.39	1.88	1.98	2.06

(i) No discount; (ii) Bundled; (iii) Exclusive; (iv) All-product

[33] Actual discount is the difference between discount price and but-for price while nominal discount is the difference between discount price and stand-alone price.

Table 2: Prices for Alternatives of Structure (a) for Setting 1

(a) Main Model

Price	# of products	Supplier S_1's View	(i)	(ii)	(iii)
$p_{A_1B_1}$	double	within-firm	3.94	3.40	3.68
$p_{A_1B_2}$	double	cross-firm	3.94	4.21	4.31
$p_{A_2B_1}$	double	cross-firm	3.94	4.21	4.31
$p_{A_2B_2}$	double	other-firm	3.94	3.89	3.84
$p_{A_1B_0}$	single	within-firm	1.97	2.27	1.84
$p_{A_2B_0}$	single	other-firm	1.97	1.95	1.92
$p_{A_0B_1}$	single	within-firm	1.97	2.27	1.84
$p_{A_0B_2}$	single	other-firm	1.97	1.95	1.92

(i) No discount; (ii) Bundled; (iii) Exclusive

Table 3: Equilibrium Prices of Structures (a), (b), and (c) for Setting 1

Price	(a) Main Model				(b) Monopoly in B				(c) Duopoly		
	(i)	(ii)	(iii)	(iv)	(i)	(ii)	(iii)	(iv)	(i)	(ii)	(iii)
$d_{A_1B_1}$	N/A	3.40	N/A	N/A	N/A	3.94	N/A	N/A	N/A	3.20	N/A
d_{A_1}	N/A	N/A	1.84	1.84	N/A	N/A	2.02	2.00	N/A	N/A	1.73
d_{B_1}	N/A	N/A	1.84	1.84	N/A	N/A	2.37	2.34	N/A	N/A	1.73
p_{A_1}	1.97	2.27	2.39	2.39	1.97	2.36	N/A	N/A	1.97	2.17	2.22
p_{B_1}	1.97	2.27	2.39	2.39	2.57	2.94	3.04	Inf[34]	1.97	2.17	2.22
$d_{A_2B_2}$	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	3.20	N/A
d_{A_2}	N/A	N/A	N/A	1.92	N/A	N/A	N/A	1.91	N/A	N/A	1.73
d_{B_2}	N/A	N/A	N/A	1.92	N/A	N/A	N/A	N/A	N/A	N/A	1.73
p_{A_2}	1.97	1.95	1.92	1.92	1.97	1.98	1.93	Inf	1.97	2.17	2.22
p_{B_2}	1.97	1.95	1.92	1.92	N/A	N/A	N/A	N/A	1.97	2.17	2.22

(i) No discount; (ii) Bundled; (iii) Exclusive; (iv) All-product

[34] Inf here denotes for prices that are effectively infinity that results in effectively 0 market share for the corresponding alternative. p_{B_1} and p_{A_2} only affects the price of alternative (A_2, B_1) in this case. As discussed in Section 3.2, both prices are set to infinity in equilibrium.

Table 4: Market Shares of Structures (a), (b), and (c) for Setting 1

Share	(a) Main Model				(b) Monopoly in B				(c) Duopoly		
	(i)	(ii)	(iii)	(iv)	(i)	(ii)	(iii)	(iv)	(i)	(ii)	(iii)
$s_{A_1B_1}$	0.09	0.21	0.14	0.14	0.11	0.25	0.15	0.18	0.09	0.22	0.15
$s_{A_1B_2}$	0.09	0.06	0.05	0.05	N/A	N/A	N/A	N/A	0.09	0.03	0.03
$s_{A_2B_1}$	0.09	0.06	0.05	0.05	0.11	0.05	0.06	0.00	0.09	0.03	0.03
$s_{A_2B_2}$	0.09	0.09	0.11	0.11	N/A	N/A	N/A	N/A	0.09	0.22	0.15
$s_{A_1B_0}$	0.12	0.08	0.13	0.13	0.19	0.12	0.18	0.18	0.12	0.09	0.13
$s_{A_2B_0}$	0.12	0.14	0.12	0.12	0.19	0.22	0.21	0.21	0.12	0.09	0.13
$s_{A_0B_1}$	0.12	0.08	0.13	0.13	0.14	0.10	0.17	0.18	0.12	0.09	0.13
$s_{A_0B_2}$	0.12	0.14	0.12	0.12	N/A	N/A	N/A	N/A	0.12	0.09	0.13

(i) No discount; (ii) Bundled; (iii) Exclusive; (iv) All-product

4.5.2 Setting 2: Small-Firm Cost Advantage

In this setting, all parameters are the same as in Setting 1 except that

$$c_{A_1} = c_{B_1} = 1.5.$$

Thus, competitors enjoy a cost advantage over Supplier S_1. Tables 5–8 provide the equilibrium under this setting. Compared to Setting 1, prices of all suppliers are higher and the profits of Supplier S_1 are lower. This is understandable as the costs of Supplier S_1's products are higher. The most important qualitative difference lies in consumer surplus for the main model (a) in Table 5, which is higher under bundled discount (1.62) than under no discount (1.61) in this setting.[35] This implies that conclusions on consumer welfare depend on market parameters. It seems that when a firm (Supplier S_1) is in a disadvantageous position (in terms of marginal cost), having the privilege of offering bundled discount increases its competitiveness and results in a more competitive market, which may benefit consumers. However, generally speaking, the degree of competition is difficult to measure

[35] In all the trials run, the differences between consumer surplus in bundled discount and no discount gets larger as I increase c_{A_1} and c_{B_1}.

when different pricing mechanisms are allowed in a multimarket competition. As discussed in Section 5.1, equilibrium prices and corresponding welfare effects are a result of balancing several difference forces. Thus, it is difficult to predict or fully understand the net effect (i.e., the equilibrium). In fact, in this example, qualitative results do not naturally extend across structures either. While consumer surplus is higher in the bundled-discount case (1.62) than in the no-discount case (1.61) under structure (a), the relation reverses under structure (b) (1.14 versus 1.15). Which price mechanism benefits the consumers is specific to the balance of the impact of discount on within-firm and cross-firm alternatives, which is determined by the market environment (preference and cost parameters) and the market structure.

Table 5: Profits and CS of Structures (a), (b), and (c) for Setting 2

Profit	(a) Main Model				(b) Monopoly in B				(c) Duopoly		
Firm	(i)	(ii)	(iii)	(iv)	(i)	(ii)	(iii)	(iv)	(i)	(ii)	(iii)
S_1	0.37	0.39	0.37	0.37	0.59	0.63	0.59	0.55	0.37	0.33	0.28
S_2	0.36	0.35	0.34	0.34	0.36	0.34	0.33	0.26	0.73	0.74	0.67
S_3	0.36	0.35	0.34	0.34	N/A	N/A	N/A	N/A	N/A	N/A	N/A
CS	1.61	1.62	1.64	1.64	1.15	1.14	1.16	1.15	1.61	1.69	1.74

(i) No discount; (ii) Bundled; (iii) Exclusive; (iv) All-product

Table 6: Prices for Alternatives of Structure (a) for Setting 2

(a) Main Model

Price	# of products	Supplier S_1's View	(i)	(ii)	(iii)
$p_{A_1B_1}$	double	within-firm	4.68	4.11	4.45
$p_{A_1B_2}$	double	cross-firm	4.42	4.58	4.66
$p_{A_2B_1}$	double	cross-firm	4.42	4.58	4.66
$p_{A_2B_2}$	double	other-firm	4.16	4.11	4.10
$p_{A_1B_0}$	single	within-firm	2.34	2.52	2.23
$p_{A_2B_0}$	single	other-firm	2.08	2.06	2.05
$p_{A_0B_1}$	single	within-firm	2.34	2.52	2.23
$p_{A_0B_2}$	single	other-firm	2.08	2.06	2.05

(i) No discount; (ii) Bundled; (iii) Exclusive

Table 7: Equilibrium Prices of Structures (a), (b), and (c) for Setting 2

Price	(a) Main Model				(b) Monopoly in B				(c) Duopoly		
	(i)	(ii)	(iii)	(iv)	(i)	(ii)	(iii)	(iv)	(i)	(ii)	(iii)
$d_{A_1B_1}$	N/A	4.11	N/A	N/A	N/A	4.63	N/A	N/A	N/A	3.95	N/A
d_{A_1}	N/A	N/A	2.23	2.23	N/A	N/A	2.37	2.33	N/A	N/A	2.14
d_{B_1}	N/A	N/A	2.23	2.23	N/A	N/A	2.74	2.70	N/A	N/A	2.14
p_{A_1}	2.34	2.52	2.61	2.61	2.34	2.61	N/A	N/A	2.34	2.47	2.47
p_{B_1}	2.34	2.52	2.61	2.61	2.90	3.13	3.24	Inf	2.34	2.47	2.47
$d_{A_2B_2}$	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	3.47	N/A
d_{A_2}	N/A	N/A	N/A	2.04	N/A	N/A	N/A	2.01	N/A	N/A	1.90
d_{B_2}	N/A	N/A	N/A	2.04	N/A	N/A	N/A	N/A	N/A	N/A	1.90
p_{A_2}	2.08	2.06	2.05	2.05	2.08	2.08	2.04	Inf	2.08	2.37	2.52
p_{B_2}	2.08	2.06	2.05	2.05	N/A	N/A	N/A	N/A	2.08	2.37	2.52

(i) No discount; (ii) Bundled; (iii) Exclusive; (iv) All-product

Table 8: Market Shares of Structures (a), (b), and (c) for Setting 2

Share	(a) Main Model				(b) Monopoly in B				(c) Duopoly		
	(i)	(ii)	(iii)	(iv)	(i)	(ii)	(iii)	(iv)	(i)	(ii)	(iii)
$s_{A_1B_1}$	0.05	0.11	0.07	0.07	0.06	0.15	0.08	0.11	0.05	0.12	0.08
$s_{A_1B_2}$	0.07	0.05	0.05	0.05	N/A	N/A	N/A	N/A	0.07	0.03	0.02
$s_{A_2B_1}$	0.07	0.05	0.05	0.05	0.10	0.06	0.06	0.00	0.07	0.03	0.02
$s_{A_2B_2}$	0.11	0.11	0.13	0.13	N/A	N/A	N/A	N/A	0.11	0.26	0.17
$s_{A_1B_0}$	0.10	0.07	0.11	0.11	0.16	0.11	0.14	0.15	0.10	0.09	0.11
$s_{A_2B_0}$	0.15	0.16	0.15	0.15	0.24	0.26	0.25	0.26	0.15	0.10	0.16
$s_{A_0B_1}$	0.10	0.07	0.11	0.11	0.13	0.10	0.15	0.16	0.10	0.09	0.11
$s_{A_0B_2}$	0.15	0.16	0.15	0.15	N/A	N/A	N/A	N/A	0.15	0.10	0.16

(i) No discount; (ii) Bundled; (iii) Exclusive; (iv) All-product

4.5.3 Setting 3: Product Advantage of Supplier S_1

In this setting, all parameters are the same as in Setting 1 except that

$$v_{ij} = \begin{cases} 8 & \text{if } i = j = 1 \\ 6 & \text{if } (i = 1, j > 1) \;\; or \;\; (i > 1, j = 1) \\ 4 & \text{if } i > 1, j > 1 \\ 2 & \text{if } i = 0 \; or \; j = 0. \end{cases}$$

Namely, for the bundled product, the bundles offered solely by Supplier S_1 generate the highest mean price-excluded utility. The bundle consisting of one product of Supplier S_1 provides higher mean price-excluded utility than those offered by other suppliers. Finally, bundled products generate higher mean price-excluded utility than a single product as in the previous settings.

Tables 9–12 provide the equilibrium under this setting. The impact of different price mechanisms is again different from Settings 1 and 2. Other than the duopoly setting, consumer surplus is highest under the no-discount case and lowest under the all-product exclusive-discount case. Taking the main model (a) as an example, consumer surplus under no discount, bundled discount, exclusive discount, and all-product exclusive discount is, respectively, 2.89, 2.59, 2.65, and 2.47. In fact, Table 10 shows that prices are the lowest for all alternatives under the no-discount case. Table 11 suggests that the exclusive-discount price (2.83) is actually higher than the but-for price (2.77). Thus, it is possible that the discount is only nominal. It seems that when a firm (Supplier S_1) is in an advantageous position (in terms of product quality), having the privilege of offering bundled discount strengthens its dominance and makes competitive constraints less effective, which hurts consumers.

Lastly, even for the duopoly structure (c), consumers are still better off without any discount than with the exclusive discount (2.84 versus 2.83). Thus, additional competition pressure with more than one firm offering discount does not necessarily dominate incentives in raising prices to exploit product advantage. This observation indicates that the belief that "exclusive discount is not anticompetitive if other firms can also offer exclusive discount" is not always valid, especially when the accused firm already has product advantages.

Table 9: Profits and CS of Structures (a), (b), and (c) for Setting 3

Profit	(a) Main Model				(b) Monopoly in B				(c) Duopoly		
Firm	(i)	(ii)	(iii)	(iv)	(i)	(ii)	(iii)	(iv)	(i)	(ii)	(iii)
S_1	2.43	2.77	2.72	2.84	3.02	3.19	3.12	3.12	2.48	2.57	2.56
S_2	0.15	0.10	0.09	0.08	0.13	0.12	0.11	0.11	0.30	0.18	0.15
S_3	0.15	0.10	0.09	0.08	N/A	N/A	N/A	N/A	N/A	N/A	N/A
CS	2.89	2.59	2.65	2.47	2.24	2.10	2.15	2.15	2.84	2.88	2.83

(i) No discount; (ii) Bundled; (iii) Exclusive; (iv) All-product

Table 10: Prices for Alternatives of Structure (a) for Setting 3

(a) Main Model

Price	# of products	Supplier S_1's View	(i)	(ii)	(iii)
$p_{A_1B_1}$	double	within-firm	5.54	5.70	5.66
$p_{A_1B_2}$	double	cross-firm	4.58	6.46	6.52
$p_{A_2B_1}$	double	cross-firm	4.58	6.46	6.52
$p_{A_2B_2}$	double	other-firm	3.63	3.88	3.73
$p_{A_1B_0}$	single	within-firm	2.77	4.52	2.83
$p_{A_2B_0}$	single	other-firm	1.81	1.94	1.87
$p_{A_0B_1}$	single	within-firm	2.77	4.52	2.83
$p_{A_0B_2}$	single	other-firm	1.81	1.94	1.87

(i) No discount; (ii) Bundled; (iii) Exclusive

Table 11: Equilibrium Prices of Structures (a), (b), and (c) for Setting 3

Price	(a) Main Model				(b) Monopoly in B				(c) Duopoly		
	(i)	(ii)	(iii)	(iv)	(i)	(ii)	(iii)	(iv)	(i)	(ii)	(iii)
$d_{A_1B_1}$	N/A	5.70	N/A	N/A	N/A	6.19	N/A	N/A	N/A	5.37	N/A
d_{A_1}	N/A	N/A	2.83	2.93	N/A	N/A	2.73	2.73	N/A	N/A	2.74
d_{B_1}	N/A	N/A	2.83	2.93	N/A	N/A	3.45	3.45	N/A	N/A	2.74
p_{A_1}	2.77	4.52	4.65	Inf	2.27	3.59	N/A	N/A	2.79	4.13	Inf
p_{B_1}	2.77	4.52	4.65	Inf	3.86	5.19	Inf	Inf	2.79	4.13	Inf
$d_{A_2B_2}$	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	2.74	N/A
d_{A_2}	N/A	N/A	N/A	1.97	N/A	N/A	N/A	1.96	N/A	N/A	1.65
d_{B_2}	N/A	N/A	N/A	1.97	N/A	N/A	N/A	N/A	N/A	N/A	1.65
p_{A_2}	1.81	1.94	1.87	Inf	1.84	2.05	1.96	Inf	1.87	2.08	Inf
p_{B_2}	1.81	1.94	1.87	Inf	N/A	N/A	N/A	N/A	1.87	2.08	Inf

(i) No discount; (ii) Bundled; (iii) Exclusive; (iv) All-product

Table 12: Market Shares of Structures (a), (b), and (c) for Setting 3

Share	(a) Main Model				(b) Monopoly in B				(c) Duopoly		
	(i)	(ii)	(iii)	(iv)	(i)	(ii)	(iii)	(iv)	(i)	(ii)	(iii)
$s_{A_1B_1}$	0.58	0.73	0.72	0.72	0.67	0.75	0.72	0.72	0.58	0.75	0.72
$s_{A_1B_2}$	0.10	0.01	0.01	0.00	N/A	N/A	N/A	N/A	0.10	0.01	0.00
$s_{A_2B_1}$	0.10	0.01	0.01	0.00	0.05	0.00	0.00	0.00	0.10	0.01	0.00
$s_{A_2B_2}$	0.02	0.02	0.02	0.00	N/A	N/A	N/A	N/A	0.02	0.08	0.03
$s_{A_1B_0}$	0.01	0.00	0.02	0.02	0.05	0.01	0.03	0.03	0.01	0.00	0.01
$s_{A_2B_0}$	0.06	0.08	0.07	0.08	0.11	0.11	0.11	0.11	0.06	0.05	0.08
$s_{A_0B_1}$	0.01	0.00	0.02	0.02	0.02	0.01	0.03	0.03	0.01	0.00	0.01
$s_{A_0B_2}$	0.06	0.08	0.07	0.08	N/A	N/A	N/A	N/A	0.06	0.05	0.08

(i) No discount; (ii) Bundled; (iii) Exclusive; (iv) All-product

4.5.4 Comments Across Parameter Settings

Comparing equilibria under different parameter settings suggests that price and welfare effects are specific to preference and cost parameters and to market structures. Discount prices could be higher or lower than but-for prices. Stand-alone prices may be set to infinity (pure exclusion) or not. Consumer surplus under the no-discount case, the bundled-discount case, and the exclusive-discount case could be in any order.[36] For fixed percentage or value discount, the impact on prices and consumer welfare compared to the free exclusive-discount case is also uncertain.[37] In summary, the impact of bundled or exclusive discount on consumer welfare is theoretically ambiguous even without considering exit and entry. The discounts also affect exit and entry differently in the model, as reflected by different equilibrium profits of competitors across settings, which are absent of any fixed or setup cost.

[36] Many additional parameter settings are evaluated, strengthening the parameter-specific conclusion. These additional results are available upon request.

[37] For the chosen parameter settings, the differences are generally not visible when we rounded the numbers in the tables. However, the order is in fact different across the settings.

4.6 Conclusion

This paper attempts to fill the gap between economic theory and antitrust practice for exclusive discount. It builds an empirical framework rooted in rigorous current economic theory and targeting consumer welfare. The framework is general in the sense that it allows heterogeneous products and bundled products. For bundled products, the idea is to treat bundles as alternatives in a discrete-choice model for demand estimation. The difference between products and alternatives complicates solving for the counterfactual equilibria. The paper then provides formulas and numerical examples to illustrate application of the framework. The numerical example suggests that the price, profit, and consumer-welfare effects of exclusive discount are case specific. Thus, instead of making rule-of-thumb judgments to classify exclusive discount as *per se* legal or illegal, the empirical framework should instead be applied in case studies. The numerical example also demonstrates that bundled and exclusive discount might have different qualitative effects on consumers. Thus, it is important to classify the type of discount in practice.

The paper should be viewed as a first step in building a rigorous empirical framework for exclusive discount and needs to be expanded to account for more general situations. One important extension is to allow a vertical structure with intermediate distributors. With the vertical structure, the partial market-share or volume discount can be more appropriately evaluated. Modeling market-share or volume discount on distributors' bulk purchases also permits multiunit purchases, which is important for some cases. Another direction of extension is to create a two-period or even dynamic version of the model to incorporate dynamic decisions. Particularly, it would be useful to allow endogenous entry and exit in evaluating the foreclosure effect of exclusive discount.

References

[1] Abito, Jose Miguel, and Julian Wright. "Exclusive dealing with imperfect downstream competition." International Journal of Industrial Organization 26.1 (2008): 227–246.

[2] Anderson, Simon P., and Luc Leruth. "Why firms may prefer not to price discriminate via mixed bundling." International Journal of Industrial Organization 11.1 (1993): 49–61.

[3] Berry, Steven T. "Estimating discrete-choice models of product differentiation." The RAND Journal of Economics 25.2 (1994): 242–262.

[4] Berry, Steven, James Levinsohn, and Ariel Pakes. "Automobile prices in market equilibrium." Econometrica 63.4 (1995): 841–890.

[5] Bierlaire, Michel. "A theoretical analysis of the cross-nested logit model." Annals of operations research 144.1 (2006): 287–300.

[6] Chen, Yongmin, and Michael H. Riordan. "Vertical integration, exclusive dealing, and expost cartelization." The Rand Journal of Economics 38.1 (2007): 1–21.

[7] Elhauge, Einer. "How loyalty discounts can perversely discourage discounting." Journal of Competition Law and Economics 5.2 (2009a): 189–231.

[8] Elhauge, Einer. "Tying, bundled discounts, and the death of the single monopoly profit theory." Harvard Law Review 123.2 (2009b): 397–481.

[9] Fumagalli, Chiara, and Massimo Motta. "Exclusive dealing and entry, when buyers compete." American Economic Review 96.3 (2006): 785–795.

[10] Fumagalli, Chiara, Massimo Motta, and Thomas Ronde. "Exclusive dealing: The interaction between foreclosure and investment promotion." working paper, 2009.

[11] Gilbert, Richard J. "Exclusive dealing, preferential dealing, and dynamic efficiency." Review of Industrial Organization 16.2 (2000): 167–184.

[12] Greenlee, Patrick, and David Reitman. "Distinguishing competitive and exclusionary uses of loyalty discounts." Antitrust Bulletin 50 (2005): 441.

[13] Greenlee, Patrick, David Reitman, and David S. Sibley. "An antitrust analysis of bundled loyalty discounts." International Journal of Industrial Organization 26.5 (2008): 1132–1152.

[14] Klein, Benjamin, and Andres Lerner. "The expanded economics of free-riding: How exclusive dealing prevents free-riding and creates undivided loyalty." Antitrust Law Journal 74.2 (2007): 473–519.

[15] Klein, Benjamin, and Kevin M. Murphy. "Exclusive dealing intensifies competition for distribution." Antitrust Law Journal 75.2 (2008): 433–466.

[16] Kobayashi, Bruce H. "The economics of loyalty discounts and antitrust law in the United States." George Mason University School of Law, 2005a.

[17] Kobayashi, Bruce H. "Two tales of bundling: Implications for the application of antitrust law to bundled discounts." George Mason University School of Law, 2005b.

[18] Kolay, S., Shaffer, G. and Ordover, J. A. "All-unit discounts in retail contracts." Journal of Economics and Management Strategy, 13 (2004): 429–459.

[19] McFadden, Daniel. "Modelling the choice of residential location." Institute of Transportation Studies, University of California, 1978.

[20] Nalebuff, Barry. "Bundling as an entry barrier." The Quarterly Journal of Economics 119.1 (2004): 159–187.

[21] Nalebuff, Barry. "Exclusionary bundling." Antitrust Bulletin 50 (2005): 321.

[22] Nalebuff, Barry. "Bundling as a way to leverage monopoly." Yale School of Management, 2008.

[23] Nevo, Aviv. "Empirical models of consumer behavior." Annual Review of Economics 3.1 (2011): 51–75.

[24] Pereira, Pedro, Tiago Ribeiro, and João Vareda. "Delineating markets for bundles with consumer level data: The case of triple-play." working paper, 2012

[25] Rey, Patrick, and Jean Tirole. "Cooperation vs. collusion: How essentiality shapes co-opetition" working paper, 2013.

[26] Steuer, Richard M. "Customer-instigated exclusive dealing." Antitrust Law Journal 68.2 (2000): 239–251.

[27] Tom, Willard K., David A. Balto, and Neil W. Averitt. "Anticompetitive aspects of market-share discounts and other incentives to exclusive dealing." Antitrust Law Journal 67.3 (2000): 615–639.

[28] Train, Kenneth. Discrete choice methods with simulation. Cambridge university press, 2003.

[29] Wright, Julian. "Exclusive dealing and entry, when buyers compete: Comment." The American Economic Review 99.3 (2009): 1070–1081.

4.7 Appendix

4.7.1 Computational Details

4.7.1.1 Alternatives and Price-Choice Variables

For the main model in Section 3, let choice alternatives be ordered as

$$(A_1, B_1), (A_1, B_2), (A_2, B_1), (A_2, B_2), (A_1, B_0), (A_2, B_0), (A_0, B_1), (A_0, B_2),$$

and order price variables as

$$\begin{aligned} \text{No discount:} \quad & (p_{A_1}, p_{B_1}, p_{A_2}, p_{B_2}) \\ \text{Bundled discount:} \quad & (d_{A_1B_1}, p_{A_1}, p_{B_1}, p_{A_2}, p_{B_2}) \\ \text{Exclusive discount:} \quad & (d_{A_1}, d_{B_1}, p_{A_1}, p_{B_1}, p_{A_2}, p_{B_2}) \\ \text{All-product exclusive discount:} \quad & (d_{A_1}, d_{B_1}, p_{A_1}, p_{B_1}, d_{A_2}, p_{A_2}, d_{B_2}, p_{B_2}). \end{aligned}$$

The firm-by-product ownership structure is

$$Ownership = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

For the monopoly-in-one-market variation, order alternatives as

$$(A_1, B_1), (A_2, B_1), (A_1, B_0), (A_2, B_0), (A_0, B_1),$$

and order price variables as

$$\begin{aligned} \text{No discount:} \quad & (p_{A_1}, p_{B_1}, p_{A_2}) \\ \text{Bundled discount:} \quad & (d_{A_1B_1}, p_{A_1}, p_{B_1}, p_{A_2}) \\ \text{Exclusive discount:} \quad & (d_{A_1}, d_{B_1}, p_{B_1}, p_{A_2}) \\ \text{All-product exclusive discount:} \quad & (d_{A_1}, d_{B_1}, p_{B_1}, d_{A_2}, p_{A_2}) \end{aligned}$$

The firm-by-product ownership structure is

$$Ownership = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \end{bmatrix}.$$

For the duopoly variation model, order alternatives as

$$(A_1, B_1), (A_1, B_2), (A_2, B_1), (A_2, B_2), (A_1, B_0), (A_2, B_0), (A_0, B_1), (A_0, B_2),$$

and order price variables as

$$\begin{aligned} \text{No discount:} &\quad (p_{A_1}, p_{B_1}, p_{A_2}, p_{B_2}) \\ \text{Bundled discount:} &\quad (d_{A_1B_1}, p_{A_1}, p_{B_1}, d_{A_2B_2}, p_{A_2}, p_{B_2}) \\ \text{Exclusive discount:} &\quad (d_{A_1}, d_{B_1}, p_{A_1}, p_{B_1}, d_{A_2}, d_{B_2}, p_{A_2}, p_{B_2}) \end{aligned}$$

The firm-by-product ownership structure is

$$Ownership = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}.$$

For the commitment-in-discount model, order alternatives as

$$\begin{aligned} &(A_1, B_1), (A_1, B_2), (A_2, B_1), (A_2, B_2), (A_3, B_1), (A_3, B_2), \\ &(A_1, B_0), (A_2, B_0), (A_3, B_0), (A_0, B_1), (A_0, B_2), \end{aligned}$$

and order price variables as

$$\begin{aligned} \text{No discount:} &\quad (p_{A_1}, p_{B_1}, p_{A_2}, p_{B_2}, p_{A_3}) \\ \text{Exclusive discount:} &\quad (d_{A_1}, d_{B_1}, p_{A_1}, p_{B_1}, p_{A_2}, p_{B_2}, p_{A_3}) \\ \text{Percentage discount:} &\quad (p_{A_1}, p_{B_1}, p_{A_2}, p_{B_2}, p_{A_3}) \\ \text{Value discount:} &\quad (p_{A_1}, p_{B_1}, p_{A_2}, p_{B_2}, p_{A_3}) \end{aligned}$$

The firm-by-product ownership structure is

$$Ownership = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}.$$

4.7.1.2 Computing $\frac{\partial s_n}{\partial p_r}$

The FOCs for the bundle alternative model are more complicated than the single-product alternative case for at least two reasons: First, the total price of an alternative can be a sum of two prices set by different firms. Thus, the same alternative market share can match different prices in different firms' FOCs. Second, one price can affect more than one alternative. For example, in computing partial derivatives of market share, $s_{A_1B_1}$, with respect to p_{A_1} using (2) for case (i), no discount, p_{A_1} presents in three alternatives (A_1, B_1), (A_1, B_2), and (A_1, B_0) , which for (A_1, B_1) are, respectively, the product itself, a same-group product, and a different-group product. Here I discuss how this affects the computation.

For alternative n in group G_k, choice probability s_n under the nested logit model is

$$s_n = \frac{e^{\delta_n/\lambda_k} \left(\sum_{m\in G_k} e^{\delta_m/\lambda_k}\right)^{\lambda_k - 1}}{\sum_{l=1}^{K} \left(\sum_{m\in G_l} e^{\delta_m/\lambda_l}\right)^{\lambda_l}}.$$

Define

$$D_k = \sum_{m\in G_k} e^{\delta_m/\lambda_k}$$

and

$$D_{all} = \sum_{l=1}^{K} \left(\sum_{m\in G_l} e^{\delta_m/\lambda_l}\right)^{\lambda_l}.$$

Suppose product r is *only* part of alternative m. Then the partial derivative of s_n of group

k with respect to p_r is

$$\frac{\partial s_n}{\partial p_r} = \begin{cases} \Delta^1_{nm} = \alpha s_n s_m \left(1 + \frac{1-\lambda_k}{\lambda_k} D_k^{-\lambda_k} D_{all}\right) - \frac{\alpha}{\lambda_k} s_n & \text{if } r \text{ is part of alternative } n \\ \Delta^2_{nm} = \alpha s_n s_m \left(1 + \frac{1-\lambda_k}{\lambda_k} D_k^{-\lambda_k} D_{all}\right) & \text{if } r \text{ is part of alternative } m \neq n \text{ and } m \in G_k \\ \Delta^3_{nm} = \alpha s_n s_m & \text{if } r \text{ is part of alternative } m \text{ and } m \notin G_k \end{cases}$$

However, as one product might be part of more than one alternative, the partial derivative is more complicated. For example, for the structure in the main model in Section 3, the matrix for partial derivatives of market shares with respect to prices for the bundled-discount case is

$$\Delta = \begin{bmatrix} \Delta^1_{11} & \Delta^2_{12} + \Delta^3_{15} & \Delta^2_{13} + \Delta^3_{17} & \Delta^2_{13} + \Delta^2_{14} + \Delta^3_{16} & \Delta^2_{12} + \Delta^2_{14} + \Delta^3_{18} \\ \Delta^2_{21} & \Delta^1_{22} + \Delta^3_{25} & \Delta^2_{23} + \Delta^3_{27} & \Delta^2_{23} + \Delta^2_{24} + \Delta^3_{26} & \Delta^1_{22} + \Delta^2_{24} + \Delta^3_{28} \\ \Delta^2_{31} & \Delta^2_{32} + \Delta^3_{35} & \Delta^1_{33} + \Delta^3_{17} & \Delta^1_{33} + \Delta^2_{34} + \Delta^3_{36} & \Delta^2_{32} + \Delta^2_{34} + \Delta^3_{38} \\ \Delta^2_{41} & \Delta^2_{42} + \Delta^3_{45} & \Delta^2_{43} + \Delta^3_{47} & \Delta^2_{43} + \Delta^1_{44} + \Delta^3_{46} & \Delta^2_{42} + \Delta^1_{44} + \Delta^3_{18} \\ \Delta^3_{51} & \Delta^3_{52} + \Delta^1_{55} & \Delta^3_{53} + \Delta^3_{57} & \Delta^3_{53} + \Delta^3_{54} + \Delta^2_{56} & \Delta^3_{52} + \Delta^3_{54} + \Delta^3_{58} \\ \Delta^3_{61} & \Delta^3_{62} + \Delta^2_{65} & \Delta^3_{63} + \Delta^3_{67} & \Delta^3_{63} + \Delta^3_{64} + \Delta^1_{66} & \Delta^3_{62} + \Delta^3_{64} + \Delta^3_{68} \\ \Delta^3_{71} & \Delta^3_{72} + \Delta^3_{75} & \Delta^3_{73} + \Delta^1_{77} & \Delta^3_{73} + \Delta^3_{74} + \Delta^3_{76} & \Delta^3_{72} + \Delta^3_{74} + \Delta^2_{78} \\ \Delta^3_{81} & \Delta^3_{82} + \Delta^3_{85} & \Delta^3_{83} + \Delta^2_{87} & \Delta^3_{83} + \Delta^3_{84} + \Delta^3_{86} & \Delta^3_{82} + \Delta^3_{84} + \Delta^1_{88} \end{bmatrix}^T .$$

4.7.2 Numerical Examples on Commitment in Discounts and Entry

In Section 4.2, I discussed the model of commitment in discounts. In a homogeneous-product theoretical model, Elhauge (2009a) shows that commitment in discounts hurts consumers as it discourages firms from competing for free buyers. In this section, I examine whether Elhauge's conclusion generalizes to my empirical framework. In addition to the market structures considered in the numerical examples in Section 5, I examine a fourth market structure: (d) commitment in discounts described in Section 4.2 where, under market structure (a), Supplier S_4 enters market A. Product A_3 of the new entrant, Supplier S_4, is assumed to be identical to product A_1 of Supplier S_1. four cases are considered: (i) no discount, (iii) (free) exclusive discount for Supplier S_1, (v) fixed-percentage exclusive discount for Supplier S_1, (vi) fixed-value exclusive discount for Supplier S_1. Comparing prices and consumer welfare under cases (iii), (v), and (vi) provides the impact of commitment in discounts. Applying this market structure to all three numerical examples in Section 5, I found no qualitative differences in any case. Commitment in discount does not necessarily lower consumer welfare. The results are demonstrated in Tables 13–21.

Note that this market structure can also be used to study the impact of entry when compared to market structure (a). In Setting 2, the entrant (Supplier S_4) also enjoys a cost advantage over Supplier S_1. However, the patterns of consumer surplus are not different from those of Setting 1 (compare the last row of Table 13 and Table 16). Thus, even a 33% $((1.5-1)/1.5)$ cost advantage favoring the entrant might not be crucial at all; the profit of Supplier S_4 is still lower under exclusive discount than under no discount (0.21 versus 0.23). Exclusive discount might lead to foreclosure even when the entrant has a cost advantage. (In this case, for example, if the fixed cost for the entrant Supplier S_4 is 0.22.)

Table 13: Profits and CS of Structure (d) for Setting 1

Profit	(d) S_4 Enter			
Firm	(i)	(iii)	(v)	(vi)
S_1	0.49	0.50	0.50	0.50
S_2	0.19	0.17	0.17	0.17
S_3	0.30	0.26	0.26	0.26
S_4	0.19	0.17	0.17	0.17
CS	2.13	2.16	2.16	2.16

Table 14: Equilibrium Prices of Structure (d) for Setting 1

Price	(d) S_4 Enter			
	(i)	(iii)	(v)	(vi)
d_{A_1}	N/A	1.71	1.72[38]	1.71
d_{B_1}	N/A	1.80	1.76	1.76
p_{A_1}	1.83	2.23	2.23	2.26
p_{B_1}	1.97	2.23	2.29	2.32
d_{A_2}	1.83	1.80	1.80	1.80
d_{B_2}	1.97	1.93	1.93	1.93
d_{A_3}	1.83	1.80	1.80	1.80

Table 15: Market Shares of Structure (d) for Setting 1

Share	(d) S_4 Enter			
	(i)	(iii)	(v)	(vi)
$s_{A_1B_1}$	0.07	0.11	0.12	0.12
$s_{A_1B_2}$	0.07	0.04	0.04	0.04
$s_{A_2B_1}$	0.07	0.05	0.04	0.04
$s_{A_2B_2}$	0.07	0.08	0.08	0.08
$s_{A_3B_1}$	0.07	0.05	0.04	0.04
$s_{A_3B_2}$	0.07	0.08	0.08	0.08
$s_{A_1B_0}$	0.09	0.10	0.10	0.10
$s_{A_2B_0}$	0.09	0.09	0.09	0.09
$s_{A_3B_0}$	0.09	0.09	0.09	0.09
$s_{A_0B_1}$	0.09	0.11	0.12	0.12
$s_{A_0B_2}$	0.09	0.09	0.09	0.09

[38] The discounted prices for cases (v) and (vi) are calculated using equilibrium prices and discount rate μ (or $\overline{d}$) calculated from case (iii) in structure (a). They are not choice variables in cases (v) and (vi).

Table 16: Profits and CS of Structure (d) for Setting 2

Profit	(d) S_4 Enter			
Firm	(i)	(iii)	(v)	(vi)
S_1	0.29	0.29	0.29	0.29
S_2	0.23	0.22	0.21	0.21
S_3	0.36	0.35	0.35	0.35
S_4	0.23	0.22	0.21	0.21
CS	1.90	1.92	1.92	1.92

Table 17: Equilibrium Prices of Structure (d) for Setting 2

Price	(d) S_4 Enter			
	(i)	(iii)	(v)	(vi)
d_{A_1}	N/A	2.12	2.12	2.12
d_{B_1}	N/A	2.21	2.17	2.17
p_{A_1}	2.23	2.48	2.49	2.50
p_{B_1}	2.34	2.48	2.54	2.55
d_{A_2}	1.87	1.86	1.86	1.86
d_{B_2}	2.08	2.06	2.06	2.06
d_{A_3}	1.87	1.86	1.86	1.86

Table 18: Market Shares of Structure (d) for Setting 2

Share	(d) S_4 Enter			
	(i)	(iii)	(v)	(vi)
$s_{A_1B_1}$	0.03	0.05	0.05	0.05
$s_{A_1B_2}$	0.05	0.03	0.03	0.03
$s_{A_2B_1}$	0.06	0.05	0.04	0.04
$s_{A_2B_2}$	0.09	0.09	0.09	0.09
$s_{A_3B_1}$	0.06	0.05	0.04	0.04
$s_{A_3B_2}$	0.09	0.09	0.09	0.09
$s_{A_1B_0}$	0.06	0.07	0.07	0.07
$s_{A_2B_0}$	0.12	0.11	0.11	0.11
$s_{A_3B_0}$	0.12	0.11	0.11	0.11
$s_{A_0B_1}$	0.07	0.09	0.09	0.09
$s_{A_0B_2}$	0.11	0.11	0.11	0.11

Table 19: Profits and CS of Structure (d) for Setting 3

Profit	(d) S_4 Enter			
Firm	(i)	(iii)	(v)	(vi)
S_1	2.21	2.49	2.49	2.49
S_2	0.09	0.06	0.06	0.06
S_3	0.15	0.09	0.08	0.08
S_4	0.09	0.06	0.06	0.06
CS	3.14	2.94	2.94	2.93

Table 20: Equilibrium Prices of Structure (d) for Setting 3

Price	(d) S_4 Enter			
	(i)	(iii)	(v)	(vi)
d_{A_1}	N/A	2.64	2.67	2.67
d_{B_1}	N/A	2.75	2.71	2.72
p_{A_1}	2.50	4.38	4.39	4.49
p_{B_1}	2.83	4.38	4.46	4.54
d_{A_2}	1.72	1.72	1.72	1.72
d_{B_2}	1.80	1.81	1.81	1.81
d_{A_3}	1.72	1.72	1.72	1.72

Table 21: Market Shares of Structure (d) for Setting 3

Share	(d) S_4 Enter			
	(i)	(iii)	(v)	(vi)
$s_{A_1B_1}$	0.53	0.70	0.70	0.70
$s_{A_1B_2}$	0.11	0.01	0.01	0.01
$s_{A_2B_1}$	0.07	0.01	0.01	0.01
$s_{A_2B_2}$	0.01	0.02	0.02	0.02
$s_{A_3B_1}$	0.07	0.01	0.01	0.01
$s_{A_3B_2}$	0.01	0.02	0.02	0.02
$s_{A_1B_0}$	0.01	0.01	0.01	0.01
$s_{A_2B_0}$	0.04	0.05	0.05	0.05
$s_{A_3B_0}$	0.04	0.05	0.05	0.05
$s_{A_0B_1}$	0.01	0.01	0.01	0.01
$s_{A_0B_2}$	0.05	0.06	0.06	0.06

4.7.3 Random-Coefficient Model

In this section, I provide an example to show how the framework works when the demand side is a random-coefficient model instead of a nested logit model. Suppose the coefficients on observed product characteristics are random:

$$\begin{aligned} u_{abi} &= X_a\beta_{Ai} + X_b\beta_{Bi} + \xi_{ab} - \alpha\left(p_a + p_b\right) + \varepsilon_{abi} \\ &= X_a\left(\beta_A + \beta_{\sigma_A}\omega_{Ai}\right) + X_b\left(\beta_B + \beta_{\sigma_B}\omega_{Bi}\right) + \xi_{ab} - \alpha\left(p_a + p_b\right) + \varepsilon_{abi} \\ &= v_{ab} - \alpha\left(p_a + p_b\right) + X_a\beta_{\sigma_A}\omega_{Ai} + X_b\beta_{\sigma_B}\omega_{Bi} + \varepsilon_{abi} \\ &= \delta_{ab} + X_a\beta_{\sigma_A}\omega_{Ai} + X_b\beta_{\sigma_B}\omega_{Bi} + \varepsilon_{abi}, \end{aligned}$$

where $(\omega_{Ai}, \omega_{Bi})$ are vectors or variables following an i.i.d. standard normal distribution.[39] ε_{abi} follows an i.i.d. Type 1 extreme-value distribution. If an alternative involves not buying in Market A (or B), the corresponding characteristics X_a (or X_b) are set to zero.

Then, the market share for alternative (a, b) is

$$\begin{aligned} s_{ab} &= \int \frac{e^{\delta_{ab} + X_a\beta_{\sigma_A}\omega_{Ai} + X_b\beta_{\sigma_B}\omega_{Bi}}}{\sum_{(a',b')\in AB} e^{\delta_{a'b'} + X_{a'}\beta_{\sigma_A}\omega_{Ai} + X_{b'}\beta_{\sigma_B}\omega_{Bi}}} d\Phi\left(\omega_{Ai}\right) d\Phi\left(\omega_{Bi}\right) \\ &= \int s_{abi} d\Phi\left(\omega_{Ai}\right) d\Phi\left(\omega_{Bi}\right), \end{aligned}$$

where s_{abi} is the choice probability of consumer i choosing alternative (a, b).

In the numerical examples, I use Monte Carlo integration. Specifically, draw R pairs of $(\omega^r_{Ai}, \omega^r_{Bi})$ from the standard normal distribution to simulate the market share

$$\widehat{s}_{ab} = \frac{1}{R}\sum_{r=1}^{R} \frac{e^{\delta_{ab} + X_a\beta_{\sigma_A}\omega^r_{Ai} + X_b\beta_{\sigma_B}\omega^r_{Bi}}}{\sum_{(a',b')\in AB} e^{\delta_{a'b'} + X_{a'}\beta_{\sigma_A}\omega^r_{Ai} + X_{b'}\beta_{\sigma_B}\omega^r_{Bi}}}.$$

Suppose product r is *only* part of alternative m, then the partial derivative of s_n with respect to p_r is

$$\frac{\partial s_n}{\partial p_r} = \begin{cases} \Delta^1_{nm} = -\int \alpha s_{ni}\left(1 - s_{ni}\right) d\Phi\left(\omega_{Ai}\right) d\Phi\left(\omega_{Bi}\right) & \text{if } m = n \\ \Delta^2_{nm} = \int \alpha s_{ni} s_{mi} d\Phi\left(\omega_{Ai}\right) d\Phi\left(\omega_{Bi}\right) & \text{if } m \neq n. \end{cases}$$

However, as any one product might be part of more than one alternative, the partial derivative is more complicated. For example, for the structure in the main model in Section 3, the

[39] When data is available, one can also allow empirical distribution for the random coefficients as in the literature.

matrix for partial derivatives of market shares with respect to prices for case (ii), bundled discount, is

$$\Delta = \begin{bmatrix} \Delta_{11}^1 & \Delta_{12}^2 + \Delta_{15}^2 & \Delta_{13}^2 + \Delta_{17}^2 & \Delta_{13}^2 + \Delta_{14}^2 + \Delta_{16}^2 & \Delta_{12}^2 + \Delta_{14}^2 + \Delta_{18}^2 \\ \Delta_{21}^2 & \Delta_{22}^1 + \Delta_{25}^2 & \Delta_{23}^2 + \Delta_{27}^2 & \Delta_{23}^2 + \Delta_{24}^2 + \Delta_{26}^2 & \Delta_{22}^1 + \Delta_{24}^2 + \Delta_{28}^2 \\ \Delta_{31}^2 & \Delta_{32}^2 + \Delta_{35}^2 & \Delta_{33}^1 + \Delta_{17}^2 & \Delta_{33}^1 + \Delta_{34}^2 + \Delta_{36}^2 & \Delta_{32}^2 + \Delta_{34}^2 + \Delta_{38}^2 \\ \Delta_{41}^2 & \Delta_{42}^2 + \Delta_{45}^2 & \Delta_{43}^2 + \Delta_{47}^2 & \Delta_{43}^2 + \Delta_{44}^1 + \Delta_{46}^2 & \Delta_{42}^2 + \Delta_{44}^1 + \Delta_{18}^2 \\ \Delta_{51}^2 & \Delta_{52}^2 + \Delta_{55}^1 & \Delta_{53}^2 + \Delta_{57}^2 & \Delta_{53}^2 + \Delta_{54}^2 + \Delta_{56}^2 & \Delta_{52}^2 + \Delta_{54}^2 + \Delta_{58}^2 \\ \Delta_{61}^2 & \Delta_{62}^2 + \Delta_{65}^2 & \Delta_{63}^2 + \Delta_{67}^2 & \Delta_{63}^2 + \Delta_{64}^2 + \Delta_{66}^1 & \Delta_{62}^2 + \Delta_{64}^2 + \Delta_{68}^2 \\ \Delta_{71}^2 & \Delta_{72}^2 + \Delta_{75}^2 & \Delta_{73}^2 + \Delta_{77}^1 & \Delta_{73}^2 + \Delta_{74}^2 + \Delta_{76}^2 & \Delta_{72}^2 + \Delta_{74}^2 + \Delta_{78}^2 \\ \Delta_{81}^2 & \Delta_{82}^2 + \Delta_{85}^2 & \Delta_{83}^2 + \Delta_{87}^2 & \Delta_{83}^2 + \Delta_{84}^2 + \Delta_{86}^2 & \Delta_{82}^2 + \Delta_{84}^2 + \Delta_{88}^1 \end{bmatrix}^T$$

In the numerical example, as for the nested logit model, I assume that product characteristics are identical for all products in markets A and B with mean price-excluded utilities, v_{ij}, being the same within a group. Specifically, I assume that

$$v_{ij} = \begin{cases} v_{AB} = 4 & \text{if } i > 0, j > 0 \\ v_A = 2 & \text{if } i > 0, j = 0 \\ v_B = 2 & \text{if } i = 0, j > 0. \end{cases}$$

In addition, assume that there is only one random coefficient for characteristics of product A or B so that ω_{Ai} and ω_{Bi} are scalar variables. Assume that demand estimation gives

$$X_a \beta_{\sigma_A} = 1;\ X_b \beta_{\sigma_B} = 1\ \forall\ (a, b) \in AB$$

Similar to Setting 1 in Section 5, the assumptions maintain as much homogeneity across products as possible. However, the equilibrium results are not directly comparable as the demand parameters are not necessarily based on the same data set. The numerical results for the main model (a) presented in Tables 22–25 below only serve as a particular example by having a random coefficient model on the demand side with $\alpha = 1$, $R = 10000$, and

$$c_{A_i} = c_{B_j} = 1\ \forall\ i, j > 0.$$

Similar observations can be drawn as for the main model of Setting 1 in Section 5.

Table 22: Profits and CS of Structure (a) for Setting 1

Profit	(a) Main Model			
Firm	(i)	(ii)	(iii)	(iv)
S_1	0.59	0.64	0.60	0.60
S_2	0.30	0.27	0.25	0.25
S_3	0.30	0.27	0.25	0.25
CS	1.88	1.87	1.92	1.92

Table 23: Prices for Alternatives of Structure (a) for Setting 1

(a) Main Model					
Price	# of products	Supplier S_1's View	(i)	(ii)	(iii)
$p_{A_1B_1}$	double	within-firm	4.91	3.99	4.62
$p_{A_1B_2}$	double	cross-firm	4.91	5.32	5.34
$p_{A_2B_1}$	double	cross-firm	4.91	5.32	5.35
$p_{A_2B_2}$	double	other-firm	4.91	4.84	4.82
$p_{A_1B_0}$	single	within-firm	2.46	2.90	2.31
$p_{A_2B_0}$	single	other-firm	2.46	2.42	2.41
$p_{A_0B_1}$	single	within-firm	2.46	2.90	2.31
$p_{A_0B_2}$	single	other-firm	2.46	2.42	2.41

Table 24: Equilibrium Prices of Structure (a) for Setting 1

(a) Main Model

Price	(i)	(ii)	(iii)	(iv)
$d_{A_1B_1}$	N/A	3.99	N/A	N/A
d_{A_1}	N/A	N/A	2.31	2.31
d_{B_1}	N/A	N/A	2.31	2.31
p_{A_1}	2.46	2.90	2.94	2.94
p_{B_1}	2.46	2.90	2.94	2.94
$d_{A_2B_2}$	N/A	N/A	N/A	N/A
d_{A_2}	N/A	N/A	N/A	2.42
d_{B_2}	N/A	N/A	N/A	2.42
p_{A_2}	2.46	2.42	2.41	2.41
p_{B_2}	2.46	2.42	2.41	2.41

Table 25: Market Shares of Structure (a) for Setting 1

(a) Main Model

Share	(i)	(ii)	(iii)	(iv)
$s_{A_1B_1}$	0.08	0.19	0.10	0.10
$s_{A_1B_2}$	0.08	0.05	0.05	0.05
$s_{A_2B_1}$	0.08	0.05	0.05	0.05
$s_{A_2B_2}$	0.08	0.08	0.08	0.08
$s_{A_1B_0}$	0.12	0.08	0.14	0.14
$s_{A_2B_0}$	0.12	0.13	0.13	0.13
$s_{A_0B_1}$	0.12	0.08	0.14	0.14
$s_{A_0B_2}$	0.12	0.13	0.13	0.12

Curriculum Vitae

Wei Zhao was born in Beijing, China on September 28, 1982. He obtained a B.A. in Finance from Renmin University of China, Beijing, China in 2005 and an M.A. in Quantitative Methods in the Social Sciences from Columbia University, New York, NY in 2007. He entered the Ph.D. program in Economics at the Johns Hopkins University in 2007.